

ENTWICKLUNG, IMPLEMENTIERUNG UND  
BEISPIELHAFTE ANWENDUNG EINES NEUARTIGEN  
PROBABILISTISCHEN BEVÖLKERUNGSPROGNOSEMODELLS

Dissertation

zur

Erlangung des akademischen Grades

doctor rerum politicarum (Dr. rer. pol.)

der Wirtschafts- und Sozialwissenschaftlichen Fakultät

der Universität Rostock

vorgelegt von

CHRISTINA BOHK, geb. am 29. Januar 1982 in Rostock

Rostock, d. 03. Mai 2011

urn:nbn:de:gbv:28-diss2011-0155-3

Datum der Verteidigung: 11. Oktober 2011

Erstgutachter: Prof. Dr. Roland Rau<sup>1</sup>

Zweitgutachter: Prof. Dr. Trifon Missov<sup>1,2</sup>

---

<sup>1</sup>Universität Rostock, Institut für Soziologie und Demographie

<sup>2</sup>Max-Planck-Institut für demografische Forschung, Rostock

## DANKSAGUNG

Ich freue mich, diese Arbeit fertig gestellt zu haben. Insgesamt hat dieser Prozess etwa 6 Jahre gedauert. In dieser Zeit haben mich etliche Personen unterstützt, bei denen ich mich an dieser Stelle aufrichtig bedanken möchte:

Meinem Betreuer Prof. Roland Rau danke ich für den neuen Boden unter meinen Füßen...; aus den zahlreichen Gesprächen habe ich immer wieder neue Energie und Anregungen für die Weiterentwicklung meiner Arbeit ziehen können.

Prof. Adelinde M. Uhrmacher und Dr. Roland Ewald (vom Lehrstuhl für Modellierung und Simulation der Universität Rostock) danke ich für die langjährige Kooperation, durch die ich meine theoretischen Überlegungen in die Praxis umsetzen konnte.

Einzelnen Fakultätsmitgliedern, insbesondere Prof. Heike Trappe und Prof. Matthias Junge, danke ich für ihre umsichtige Unterstützung. Darüber hinaus danke ich Astrid Jahncke für die aufmunternden Gespräche.

Dr. Nadja Milewski, Dr. Jutta Gampe und Thomas Salzmann danke ich für konstruktive und aufschlussreiche Diskussionen über einzelne Aspekte meiner Arbeit.

Dr. Katja Köppen und Matthias Leuchter danke ich für die überaus unterhaltsamen (Mittags)Gespräche, die oftmals durch Kaffee und Scho-Co-Ki versüßt worden sind — ein besseres Auftanken während der Arbeit kann ich mir kaum vorstellen.

Schließlich danke ich meiner Familie für ihr Verständnis, ihre unablässige Ermunterung und ihre wertvollen Ratschläge während dieser Zeit.

Rostock, April 2011

# Inhaltsverzeichnis

<b>1. Einleitung</b>	<b>10</b>
<b>I. Theoretische Aspekte aus dem Bereich der Bevölkerungsprognosen</b>	<b>15</b>
<b>2. Terminologie und Ablauf von Bevölkerungsprognosen</b>	<b>16</b>
2.1. Grundlegende Terminologie von Bevölkerungsprognosen . . . . .	16
2.2. Genereller Ablauf eines Bevölkerungsprognoseprozesses . . . . .	18
<b>3. Entwicklung der Bevölkerungsprognose von den ersten Anfängen bis zur Gegenwart</b>	<b>22</b>
3.1. Bevölkerungsprognosen in der Vergangenheit . . . . .	23
3.1.1. Erste Ansätze zur Bevölkerungsprognose . . . . .	23
3.1.2. Erste moderne Ansätze zur Bevölkerungsprognose . . . . .	27
3.1.3. Übergang von den Bevölkerungsprognosen aus der Vergangenheit in die Gegenwart . . . . .	31
3.2. Bevölkerungsprognosen in der Gegenwart . . . . .	31
3.2.1. Kohorten-Komponenten Methode . . . . .	31
3.2.2. Deterministische und probabilistische Ansätze zur Bevölkerungsprognose . . . . .	41
3.2.3. Mikromodelle und Makromodelle zur Bevölkerungsprognose . . . . .	45
3.2.4. Strukturmodelle zur Bevölkerungsprognose . . . . .	47
3.2.5. <i>Projections</i> und <i>Forecasts</i> . . . . .	49
3.2.6. Zusammenfassung . . . . .	50
<b>4. Überblick über probabilistische Bevölkerungsprognosemodelle</b>	<b>51</b>
4.1. Fehler- und Unsicherheitsquellen beim Bevölkerungsprognoseprozess . . . . .	52
4.2. Systematik probabilistischer Ansätze zur Bevölkerungsprognose . . . . .	57
4.2.1. Extrapolation der totalen Bevölkerungszahl . . . . .	59

4.2.2.	Fehlerfortschreibung . . . . .	61
4.2.3.	Extrapolation der Bevölkerung nach Alter und Geschlecht unter Berücksichtigung demographischer Komponenten . . . . .	63
4.2.4.	Expertenmeinung und Simulation . . . . .	66
4.2.5.	Verbindung verschiedener Methoden . . . . .	73
4.3.	Schlussfolgerung — Anforderungen an das PPPM . . . . .	77
<b>5.</b>	<b>Generierung von Mortalitätsannahmen</b>	<b>80</b>
5.1.	Bestimmung eines geeigneten Parameters zur Mortalitätsprognose . . . . .	81
5.2.	Überblick über verschiedene Ansätze zur Mortalitätsprognose . . . . .	82
5.2.1.	Extrapolative Ansätze . . . . .	83
5.2.2.	Ansätze mit Zielverteilungen . . . . .	94
5.2.3.	Multikausale Ansätze . . . . .	98
5.2.4.	Sonstige Ansätze . . . . .	107
5.3.	Schlussfolgerung — Mortalitätsannahmen . . . . .	109
<b>6.</b>	<b>Generierung von Fertilitätsannahmen</b>	<b>112</b>
6.1.	Überblick über verschiedene Ansätze zur Fertilitätsprognose . . . . .	115
6.1.1.	Extrapolation . . . . .	115
6.1.2.	Interpolation . . . . .	121
6.1.3.	Erklärende theoretische bzw. Qualitative Ansätze . . . . .	123
6.1.4.	Methodenmix bzw. Multikausale Ansätze . . . . .	125
6.2.	Schlussfolgerung — Fertilitätsannahmen . . . . .	127
<b>7.</b>	<b>Generierung von Migrationsannahmen</b>	<b>129</b>
7.1.	Überblick über verschiedene Ansätze zur Migrationsprognose . . . . .	131
7.1.1.	Ansätze zur Prognose der Bruttomigration . . . . .	131
7.1.2.	Ansätze zur Prognose der Fertilität und Mortalität einzelner Migranten-Subpopulationen . . . . .	138
7.2.	Schlussfolgerung — Migrationsannahmen . . . . .	144
<b>8.</b>	<b>Bewertung von Bevölkerungsprognosen</b>	<b>146</b>
8.1.	Fehlermaße . . . . .	147
8.1.1.	Herkömmliche Fehlermaße für deterministische Bevölkerungsprognosen . . . . .	148
8.1.2.	Fehlermaße für probabilistische Bevölkerungsprognosen . . . . .	155

8.2.	Komplexe versus simple Prognosemodelle . . . . .	161
8.3.	Schlussfolgerung — Anforderungen an das PPPM . . . . .	163
<b>II. Probabilistic Population Projection Model (PPPM)</b>		<b>166</b>
<b>9.</b>	<b>Grundlegende Modellparameter des PPPM</b>	<b>167</b>
9.1.	Zu unterscheidende Subpopulationen im PPPM . . . . .	168
9.1.1.	Jump-off Bevölkerung . . . . .	169
9.1.2.	Einheimische . . . . .	170
9.1.3.	Immigranten . . . . .	170
9.1.4.	Emigranten . . . . .	171
9.1.5.	Gesamtbevölkerung . . . . .	171
9.1.6.	Theoretischer Hintergrund . . . . .	172
9.2.	Formale Definition der zu verwendenden Modellparameter im PPPM . . . .	173
9.2.1.	Jump-off Bevölkerung . . . . .	173
9.2.2.	Direkte Zu- und Fortzüge . . . . .	173
9.2.3.	Altersspezifische Fertilitätsraten . . . . .	173
9.2.4.	Überlebende im Alter $x$ . . . . .	174
9.2.5.	Geschlechteranteil bei Geburt . . . . .	175
9.2.6.	Anteil der Säuglingssterbefälle im ersten Halbjahr . . . . .	175
9.2.7.	Überlebenswahrscheinlichkeit der obersten, nach oben offenen Altersgruppe . . . . .	175
9.2.8.	Überblick . . . . .	176
9.3.	P3J — Dateneingabe . . . . .	176
9.3.1.	Das P3J und JAMES II . . . . .	176
9.3.2.	Interne Datenverwaltung . . . . .	179
9.3.3.	Nutzeroberfläche zur Eingabe der Daten . . . . .	180
<b>10.</b>	<b>Berechnung und Ausführung des PPPM</b>	<b>183</b>
10.1.	Berechnung <i>einer</i> Annahmenverlaufskombination bzw. eines <i>Ergebnisverlaufs</i>	185
10.1.1.	Halbjährige Überlebenswahrscheinlichkeiten . . . . .	187
10.1.2.	Jahresdurchschnittsbestände überhalb von Alter 0 . . . . .	189
10.1.3.	Geburten . . . . .	193
10.1.4.	Jahresendbestände . . . . .	195

10.1.5. Zusammenfassung . . . . .	198
10.2. Berechnung <i>mehrerer</i> Annahmenverlaufskombinationen bzw. einer Ergebnisverteilung . . . . .	199
10.2.1. Erstellung einer Annahmenverteilung für jeden Modellparameter . .	200
10.2.2. Kombination der Annahmenverläufe im Open Type und Limited Type	202
10.3. Berechnung der Ergebnisverteilung . . . . .	204
10.3.1. Kombinatorische Betrachtung — Eintrittswahrscheinlichkeiten und Anzahl der Annahmenverlaufskombinationen . . . . .	204
10.3.2. Alternative Berechnungsmodi des Open Type und Limited Type . .	207
10.4. P3J — Berechnung der Prognoseergebnisse . . . . .	215
10.4.1. Simulativer Ansatz . . . . .	216
10.4.2. (Quasi-)Vollständiger Ansatz . . . . .	219
10.4.3. Überprüfung der Eintrittswahrscheinlichkeiten . . . . .	221
10.4.4. Unterstützung von repräsentativen Annahmenverläufen . . . . .	221
<b>11. Prognoseergebnisse des PPPM</b>	<b>223</b>
11.1. Ergebnisgrößen . . . . .	224
11.2. Ergebnisreport . . . . .	226
11.2.1. Entwicklung der Ergebnisverteilung diverser Ergebnisgrößen über den Projektionshorizont . . . . .	227
11.2.2. Entwicklung der Wahrscheinlichkeitsverteilung diverser Ergebnisgrößen für ausgewählte Projektionsintervalle . . . . .	228
11.2.3. Entwicklung der Alters- und Geschlechtsstruktur diverser Ergebnisgrößen . . . . .	229
11.3. Erweiterte Ergebnisanalyse . . . . .	231
11.3.1. Miniprojektionen . . . . .	231
11.3.2. Levelplot . . . . .	234
11.4. P3J — Ausgabe und Zusammenstellung der Prognoseergebnisse . . . . .	237
11.4.1. Datenexport . . . . .	237
11.4.2. Reportgenerierung . . . . .	238
<b>12. Beispielhafte Anwendung des PPPM und P3J</b>	<b>241</b>
12.1. Fragestellung bzw. Prognosegegenstand . . . . .	241
12.2. Auswahl eines geeigneten Prognosemodells . . . . .	242
12.3. Sammlung relevanter Informationen/Daten . . . . .	242

12.4. Generierung der Annahmen für die Modellparameter . . . . .	242
12.4.1. Mortalität . . . . .	243
12.4.2. Fertilität . . . . .	253
12.4.3. Migration . . . . .	259
12.4.4. Sonstige Annahmen . . . . .	265
12.5. Berechnung und Analyse der Prognoseergebnisse . . . . .	267
12.5.1. Eingabe der Annahmenverläufe ins P3J . . . . .	268
12.5.2. Berechnung der Prognoseergebnisse im P3J . . . . .	277
<b>13. Zusammenfassung und Ausblick</b>	<b>290</b>
13.1. Motivation . . . . .	290
13.2. Anforderungen an ein <i>umfassendes</i> probabilistisches Bevölkerungsprognosemodell . . . . .	290
13.2.1. Auswahl geeigneter Modellparameter und Bestimmung ihres Detailgrads . . . . .	291
13.2.2. Generierung einer heterogenen Annahmenverteilung für jeden Modellparameter . . . . .	292
13.2.3. Bestimmung einer Ergebnisverteilung für wesentliche Ergebnisgrößen	293
13.2.4. Erzeugung und Analyse aussagekräftiger Prognoseergebnisse . . . .	294
13.3. Umsetzung dieser Anforderungen im PPPM und P3J . . . . .	294
13.3.1. Auswahl geeigneter Modellparameter und Bestimmung ihres Detailgrads . . . . .	295
13.3.2. Generierung einer heterogenen Annahmenverteilung für jeden Modellparameter . . . . .	295
13.3.3. Bestimmung einer Ergebnisverteilung für wesentliche Ergebnisgrößen	296
13.3.4. Erzeugung und Analyse aussagekräftiger Prognoseergebnisse . . . .	298
13.3.5. Prototypische Realisierung des PPPM als Software im P3J . . . . .	299
13.4. Ausblick bzw. noch umzusetzende Anforderungen im PPPM und P3J . . . .	299
13.4.1. Auswahl geeigneter Modellparameter und Bestimmung ihres Detailgrads . . . . .	300
13.4.2. Generierung einer heterogenen Annahmenverteilung für jeden Modellparameter . . . . .	301
13.4.3. Bestimmung einer Ergebnisverteilung für wesentliche Ergebnisgrößen	302
13.4.4. Realisierung der potentiellen PPPM-Weiterentwicklungen im P3J . .	303



13.5. Schlussbemerkung . . . . .	304
<b>Literaturverzeichnis</b>	<b>306</b>
<b>Abbildungsverzeichnis</b>	<b>335</b>
<b>Tabellenverzeichnis</b>	<b>339</b>
<b>A. Installation des P3J</b>	<b>340</b>
A.1. Benötigte Software . . . . .	340
A.1.1. Java . . . . .	340
A.1.2. MySQL . . . . .	340
A.1.3. R und L <sup>A</sup> T <sub>E</sub> X . . . . .	341
A.2. Benötigte Java-Komponenten . . . . .	341
A.3. Ausführung . . . . .	342
A.3.1. Vorbereitungen . . . . .	342
A.3.2. P3J starten . . . . .	344
A.4. Technische Dokumentation . . . . .	344
<b>B. Reihenfolge der demographischen Ereignisse im PPPM</b>	<b>345</b>
<b>C. <i>Repräsentative</i> Annahmenverläufe</b>	<b>351</b>

# 1. Einleitung

*You see, one thing is, I can live with doubt, and uncertainty, and not knowing. I think it's much more interesting to live not knowing than to have answers which might be wrong. I have approximate answers [...] and different degrees of certainty about different things. [...].*

Richard P. Feynman, Nobelpreisträger für Physik [86]

Bevölkerungsprognosen versuchen die zukünftige Entwicklung einer Bevölkerung zu prognostizieren — dies gelingt aber mit an Sicherheit grenzender Wahrscheinlichkeit nicht, wenn sie lediglich einen zukünftigen Entwicklungspfad angeben.

Bei einer Bevölkerungsprognose werden verfügbare Informationen wesentlicher Einflussfaktoren der Bevölkerungsentwicklung eingesetzt. Allerdings ist weder die zukünftige Entwicklung der Einflussfaktoren noch die der zu prognostizierenden Bevölkerung mit Sicherheit bekannt, sodass vielmehr das Spektrum potentieller Bevölkerungsentwicklungen aufgezeigt und quantifiziert werden sollte. Dabei wird dann anstatt *einer* konkreten Bevölkerungsentwicklung, deren Wahrscheinlichkeit, tatsächlich in der Zukunft einzutreten, ohnehin relativ gering ist, eine Verteilung mit *mehreren* potentiellen Bevölkerungsentwicklungen (mit unterschiedlichen Eintrittswahrscheinlichkeiten) prognostiziert, die dafür aber auch die tatsächlich in der Zukunft eintretende Bevölkerungsentwicklung mit einer relativ hohen Wahrscheinlichkeit enthält. Entsprechend der Aussage Richard P. Feynmans ist es demnach besser eine näherungsweise Lösung mit zugeordneten Wahrscheinlichkeiten als gar keine oder, schlimmer noch, *eine* falsche Lösung zu haben.

Die Prognose der zukünftigen Bevölkerungsentwicklung kann als Grundlage vielfältiger Entscheidungen u. a. in den Bereichen der Politik, Wirtschaft, Wissenschaft und Umwelt dienen. So kann beispielsweise die zukünftige Bevölkerung (in Zahl und Struktur (z. B. nach dem Alter und dem Geschlecht)) zur Umgestaltung sozialer Sicherungssysteme (wie

der Gesundheits-, Pflege-, Arbeitslosen und Rentenversicherung), zur Planung der in Zukunft zu benötigten Wohnungen (in Anzahl und Größe), Krankenhäuser und Schulen wie auch zur adaptiven Ausrichtung (der Produktpalette) einzelner Unternehmen auf zukünftig wachsende Kundengruppen genutzt werden. Darüber hinaus kann damit der Fokus auf Forschungsfelder neu ausgerichtet werden, sodass sowohl die Bedingungen der zukünftig wachsenden Altersgruppen (wie z.B. mit der Altersforschung im soziologischen und medizinischen Bereich) als auch die Ursachen für die zukünftig schrumpfenden Altersgruppen (wie z.B. mit der Familiendemographie) näher analysiert werden. Ferner können mit Bevölkerungsprognosen auch die bevölkerungsdynamischen Auswirkungen bestimmter Umweltbedingungen bzw. äußerer Einflüsse (wie z. B. einer Pandemie) untersucht werden.

Aufgrund der weitreichenden Bedeutung der Bevölkerungsprognosen für die zukünftige Gestaltung diverser gesellschaftlicher Lebensbereiche ist eine fortlaufende Verbesserung ihrer Methodik unabdingbar. Zumal die Überprüfung vergangener Bevölkerungsprognosen (von nationalen Statistikämtern in Europa in den letzten 60 Jahren) zeigt, dass deren Genauigkeit weder durch die bessere Datenqualität noch durch die bislang verbesserte Methodik wesentlich gesteigert werden konnte [129].

In der Demographie werden grundsätzlich je zwei verschiedene Arten der Prognose hinsichtlich ihrer eigentlichen Zielstellung und der Berücksichtigung von Unsicherheit unterschieden. Einerseits gibt es die so genannten *forecasts*, die versuchen, die tatsächlich in der Zukunft eintretende Bevölkerungsentwicklung zu prognostizieren. Allerdings ist die zukünftige Bevölkerungsentwicklung nicht *deterministisch*, sondern *probabilistisch*, sodass es anstatt nur einer *mehrere* potentielle zukünftige Bevölkerungsentwicklungen gibt, die in einem Bevölkerungsprognosemodell angemessen berücksichtigt und quantifiziert werden können sollten. Andererseits gibt es — im Gegensatz zu den *forecasts* — noch die so genannten *projections*, die versuchen, die bevölkerungsdynamischen Auswirkungen bestimmter (auch einzelner) Annahmen zu ermitteln.

Bisherige probabilistische Bevölkerungsprognosemodelle versuchen zwar bei der Prognose der tatsächlich in der Zukunft eintretenden Bevölkerungsentwicklung die verschiedenen potentiellen zukünftigen Bevölkerungsentwicklungen einzufangen. Jedoch können sie oftmals aufgrund einer ungenauen Modellierung der demographischen Komponenten (Fertilität, Mortalität und vor allem der Migration) und einer einschränkenden Annahmengenerierung

(u. a. durch die Vorgabe modell-basierter Methode(n) und einer festen Verteilungsform für deren Eintrittswahrscheinlichkeiten) das Spektrum der *wahren* Annahmen- und Ergebnisverteilung (der zu berücksichtigenden Modellparameter und Ergebnisgrößen) gerade bei besonders unsicheren Prognosen nur ungenügend abdecken, wodurch die Prognosegenauigkeit beträchtlich geschmälert werden kann.

Deshalb wird in dieser Arbeit ein neuartiges probabilistisches Bevölkerungsprognosemodell, das *Probabilistic Population Projection Model* (PPPM), theoretisch entwickelt und implementiert, das u. a. durch die genaue Modellierung demographischer Modellparameter und der flexiblen Generierung ihrer Annahmen (ohne die einschränkende Vorgabe modell-basierte(r) Methode(n) und einer Verteilungsform für deren Eintrittswahrscheinlichkeiten) das Spektrum der *wahren* Annahmen- und Ergebnisverteilung näherungsweise — sofern es die verfügbaren Informationen erlauben — abdecken kann.

Die vorliegende Arbeit gliedert sich (neben der Einleitung und der Zusammenfassung) in zwei Teile. Der erste Teil führt in die allgemeine Theorie bestehender Prognosemodelle für eine Bevölkerung wie auch für einzelne demographische Komponenten ein. Dabei werden in Kapitel 2 relevante Begriffe erläutert und der Ablauf eines Prognoseprozesses skizziert. In den darauf folgenden Kapiteln werden die einzelnen Prognoseschritte näher beleuchtet. So wird im Kapitel 3 die Entwicklung der Methodik zur Bevölkerungsprognose von den ersten Anfängen bis in die Gegenwart beschrieben. Dabei liegt der Fokus auf der (zunehmenden) Erkennung und Berücksichtigung der Prognoseunsicherheit in Bevölkerungsprognosemodellen, wobei bestehende probabilistische Bevölkerungsprognosemodelle nochmals gesondert im Kapitel 4 ausführlich dargestellt werden. Da sich probabilistische Bevölkerungsprognosemodelle mitunter in der Generierung der Annahmen(verteilung) unterscheiden, wird zudem in den Kapiteln 5, 6 und 7 jeweils ein Überblick über die bestehenden Methoden zur Generierung der Mortalitäts-, Fertilitäts- und Migrationsannahmen gegeben. Jedes dieser theoretischen Kapitel endet mit einer Zusammenfassung der bestehenden Methodik und den daraus zu ziehenden Konsequenzen für die Entwicklung des neuartigen probabilistischen Bevölkerungsprognosemodells PPPM. Abschließend werden im Kapitel 8 Methoden zur Bewertung deterministischer Bevölkerungsprognosen betrachtet und für probabilistische Bevölkerungsprognosen erweitert bzw. neu entwickelt, wobei auch grundsätzlich zu klären ist, inwieweit Bevölkerungsprognosen überhaupt hinsichtlich ihrer Genauigkeit bewertet werden können.

Der zweite Teil der Arbeit beschäftigt sich mit der methodischen Entwicklung, Implementierung und Anwendung des PPPM. In den Kapiteln 9 bis 11 werden dafür die Dateneingabe, Berechnung und Ausgabe bzw. Analyse der Prognoseergebnisse als drei wesentliche Aspekte des PPPM und dessen prototypischer Realisierung als Software, dem P3J, interdependent erklärt. Während im Kapitel 9 zunächst die separate Prognose einzelner Subpopulationen, respektive der Einheimischen, der direkten Zu- und Fortzüge und ihrer einzelnen Nachkommengenerationen, demographisch motiviert und die dafür zu benötigten Modellparameter formal erläutert werden, werden im Kapitel 10 darauf aufbauend die deterministischen Modellgleichungen zur Berechnung wesentlicher Ergebnisgrößen vorgestellt. Anschließend wird erklärt, wie (nur) im PPPM (bzw. P3J) mehrere Annahmenverläufe pro Modellparameter eingegeben, flexibel gewichtet und letztlich in mehreren Prognosedurchläufen kombiniert werden können, sodass auf der Basis des Berechnungsmodells anstatt eines deterministischen Ergebnisverlaufs eine probabilistische Ergebnisverteilung resultiert, mit der die Prognoseunsicherheit adäquat eingefangen und quantifiziert werden kann. Um dem Benutzer relativ viel Freiheit bei der Durchführung einer Bevölkerungsprognose mit dem PPPM (bzw. P3J) zu überlassen, kann grundsätzlich zwischen dem *Open Type* und dem *Limited Type* des PPPM sowie dem simulativen und (quasi-)vollständigen Berechnungsmodus gewählt werden. Der *Open Type* und der *Limited Type* unterscheiden sich z. B. darin, Annahmenverläufe eines oder/und mehrerer Modellparameter miteinander zu kombinieren, sodass unplausible Annahmenverlaufskombinationen — jeweils angepasst an die Erwartungen eines Prognostikers — gezielt eliminiert werden können. Daneben eignen sich die Berechnungsmodi für die Berechnung einer kompletten Ergebnisverteilung oder zur erweiterten Ergebnisanalyse bei einer jeweils unterschiedlich großen Anzahl an Annahmenverlaufskombinationen. Im Kapitel 11 wird dann die Ausgabe und Analyse der Prognoseergebnisse beschrieben. Eine Besonderheit resultiert dabei aus der separaten Prognose einzelner Subpopulationen: Während in herkömmlichen Bevölkerungsprognosemodellen (auf der Makroebene) häufig nur die zukünftige Entwicklung der Gesamtbevölkerung (disaggregiert nach Alter und Geschlecht) betrachtet werden kann, kann beim PPPM (bzw. P3J) zusätzlich die zukünftige Entwicklung einzelner (und variabel zusammengesetzter) Subpopulationen analysiert werden. Weitere Vorzüge ergeben sich aus der Assoziation der Prognoseergebnisse mit den sie erzeugenden Annahmen (pro Modellparameter). So kann beispielsweise die Auswirkung ausgewählter Annahmen (ausgewählter Modellparameter) auf die Höhe der Prognoseergebnisse analysiert werden. Daneben besteht auch die Möglichkeit, automatisch einen Ergebnisreport erstellen

zu lassen, der wesentliche Prognoseergebnisse graphisch zusammenfasst, sodass häufig bei einer Bevölkerungsprognose anfallende Arbeitsschritte auf ein Minimum reduziert werden können. Das PPPM und seine prototypische Realisierung P3J werden abschließend im Kapitel 12 bei einer umfassenden Beispielprojektion mit realen Daten zur Bevölkerung Deutschlands praktisch eingesetzt, um das generelle Prozedere bei der Durchführung einer probabilistischen Bevölkerungsprognose mit dem PPPM (bzw. P3J) aufzuzeigen. Der praktische Vergleich mit herkömmlichen (probabilistischen) Bevölkerungsprognose(modelle)n offenbart dabei zahlreiche konzeptionelle Vorzüge des PPPM, um letztlich die Prognoseunsicherheit angemessen einzufangen und eine Ergebnisverteilung zu generieren, die die in der Zukunft eintretende Bevölkerungsentwicklung auch tatsächlich (bzw. mit einer relativ hohen Wahrscheinlichkeit) enthält.

Aufgrund dieser umfangreichen Ausrichtung der Arbeit müssen einige inhaltliche Limitierungen vorgenommen werden. So konzentriert sich der theoretische Teil auf die vielfältige und einflussreiche internationale Literatur zu bestehenden Bevölkerungsprognosemodellen und Methoden zur Generierung von Annahmen zur Mortalität, Fertilität und Migration. Daneben liegt der Fokus dieser Arbeit auf der Entwicklung des neuartigen probabilistischen Bevölkerungsprognosemodells PPPM. Dabei geht es primär um die *konkrete* methodisch-technische Umsetzung der im Theorieteil abgeleiteten Anforderungen an ein möglichst realistisches und *umfassendes* Prognosemodell im PPPM, sodass damit letztendlich auf der Makroebene eine Bevölkerung möglichst flexibel und genau nach der Zahl und Struktur prognostiziert werden kann. Anstatt einer echten Bevölkerungsprognose (im Sinne der *forecasts*) werden ausführlich die methodisch-technischen Besonderheiten des PPPM im Vergleich zu herkömmlichen (probabilistischen) Bevölkerungsprognose(modelle)n anhand einer umfangreichen Beispiel*projektion* gezeigt und verdeutlicht.

## Teil I.

# Theoretische Aspekte aus dem Bereich der Bevölkerungsprognosen

## 2. Terminologie und Ablauf von Bevölkerungsprognosen

Die Beschäftigung mit Bevölkerungsprognosen erfordert die Kenntnis einiger grundlegender Begriffe wie auch die des formalen Aufbaus eines Prognoseprozesses.

### 2.1. Grundlegende Terminologie von Bevölkerungsprognosen

Bevölkerungsprognosen prognostizieren die zukünftige Bevölkerungsentwicklung über *Annahmen* wesentlicher Erklärungsfaktoren, die als *Modellparameter* bezeichnet werden. Oftmals wird zur Generierung der Annahmen die quantitative Entwicklung der Bevölkerung aus der Vergangenheit — aus der so genannten *Basisperiode* — herangezogen. Diese Basisperiode erstreckt sich vom *Basisjahr*, dem Jahr mit den frühesten verfügbaren Daten, bis zum *Jump-off Jahr*, dem Jahr mit den aktuell verfügbaren Daten. Der *Projektionshorizont*, abgegrenzt vom *Startjahr* und vom *End- oder Zieljahr*, umfasst den gesamten Zeitraum der Prognose. Eine Bevölkerungsprognose findet bei längeren Projektionshorizonten nicht direkt vom Start- zum Zieljahr der Prognose, sondern in einzelnen Schritten — so genannten *Projektionsintervallen* — statt. Bei einer Bevölkerungsprognose nach Alter entspricht die Länge der Projektionsintervalle der Länge der Altersintervalle oder einem Vielfachen davon [256]. In der Abbildung 2.1 sind diese grundlegenden Begriffe einer Prognose graphisch veranschaulicht.

Als *Punktprognose* werden Bevölkerungsprognosen zu einem bestimmten Zeitpunkt in der Zukunft bezeichnet. So kann eine Bevölkerungsprognose für ein Projektionsintervall oder auch für das End- oder Zieljahr einer Prognose als Punktprognose bezeichnet werden.

Im Zusammenhang mit Bevölkerungsprognosen werden häufig Begriffe aus relevanten Bereichen wie z.B. der *Systemtheorie* oder der computergestützten *Modellierung und Simulation* verwendet. So stellt eine Bevölkerung ein zu betrachtendes *System* dar, dem



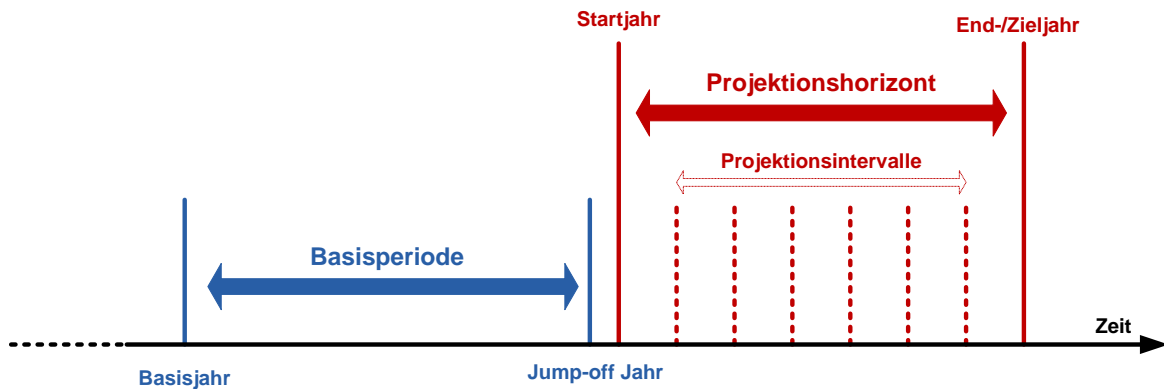


Abbildung 2.1.: Grundlegende Terminologie von Bevölkerungsprognosen

wesentliche Beobachtungsdaten entnommen werden können. Darüber hinaus werden in einem *experimentellen Rahmen* die bei einer Prognose zu berücksichtigenden Variablen bzw. Modellparameter abgegrenzt. Die meist funktionalen Zusammenhänge der auszuwählenden Modellparameter werden beispielsweise über Gleichungen im *Modellformalismus* festgelegt. Hingegen sind die vom Prognostiker zu generierenden Annahmen über die zukünftige Entwicklung der Modellparameter dem eigentlichen *Prognosemodell* zuzuordnen. Das Prognosemodell gibt demnach das Verhalten über Eingabedaten für die Modellparameter vor, um daraus entsprechende Ausgabedaten für die *Ergebnisgrößen* (wie die zukünftige Bevölkerungszahl) zu generieren. Dabei wird das Verhalten des Modells von einem *Simulator* auf der Basis des Modellformalismus berechnet bzw. mit Hilfe eines Simulationsalgorithmus auf einem Computer ausgeführt [16, 39, 321].<sup>1</sup>

Die Komplexität eines Modells wird von der Anzahl der zu berücksichtigenden Modellparameter, ihrer funktionalen Zusammenhänge wie auch von den für sie einzugebenden Annahmen bestimmt. Mit zunehmender Komplexität eines Modells wird dessen analytische Lösung zunehmend schwieriger, sodass nur noch eine *simulative* Lösung praktikabel ist. Bei der simulativen Lösung eines Modells wird das über die Annahmen zu bestimmende Modellverhalten in einem oder mehreren *Simulationsdurchläufen* simuliert, sodass anhand dieses *Simulationsexperiments* schließlich die Ausgabewerte für die Ergebnisgrößen geschätzt werden können [16, 95].

<sup>1</sup>Im weiteren Verlauf der Arbeit wird — sofern es unnötig ist — nicht weiter zwischen einem Modell und seinem zugehörigen Modellformalismus unterschieden. Anstatt dessen wird der Begriff *Modell* für beide Begriffe synonym verwendet.

Modelle können nach verschiedenen Merkmalen unterschieden werden. So können beispielsweise *diskrete* von *kontinuierlichen* Modellen oder auch *deterministische* von *probabilistischen* Modellen unterschieden werden. Bei diskreten Modellen können sich die Annahmenwerte der Modellparameter nur zu diskreten Zeitpunkten (oder endlich oft), quasi beim Übergang zwischen zwei aufeinanderfolgenden Projektionsintervallen, verändern. Im Gegensatz dazu können sich bei einem kontinuierlichen Modell die Annahmenwerte kontinuierlich, also unendlich oft, in einem Zeitintervall (wie dem Projektionshorizont) verändern.<sup>2</sup> Deterministische und probabilistische (bzw. stochastische) Bevölkerungsprognosemodelle unterscheiden sich hingegen in der Zufälligkeit der Annahmenwerte für die Modellparameter. Während deterministische Modelle Annahmenwerte für die Modellparameter für einen Prognosedurchlauf genau festlegen, sind diese bei probabilistischen Modellen (ganz oder teilweise) zufallsabhängig. Demzufolge können deterministische Modelle genau einen Ergebnisverlauf<sup>3</sup> erzeugen, wohingegen probabilistische Modelle unterschiedliche Ergebnisverläufe erzeugen können [16]. Neben den bereits aufgeführten Merkmalen können Modelle noch dahingehend unterschieden werden, ob sie die Bevölkerung auf der *Mikro*- oder auf der *Makroebene* analysieren. Dabei ziehen Mikromodelle einzelne Individuen und Makromodelle die gesamte Bevölkerung als Untersuchungsgegenstand heran.<sup>4</sup>

### 2.2. Genereller Ablauf eines Bevölkerungsprognoseprozesses

Der Ablauf einer Bevölkerungsprognose kann in Anlehnung an Armstrong [12, 13], Cruijssen und Keilman [66] und Willekens [307] in verschiedene Arbeitsschritte unterteilt werden, die in Abbildung 2.2 veranschaulicht sind.

Zuerst muss die Frage- bzw. Zielstellung der auszuführenden Prognose genau festgelegt werden. Dazu werden der Prognosegegenstand und dessen Abstraktionsniveau wie auch die Länge des Projektionshorizonts bestimmt. Der Prognosegegenstand bezeichnet die Ergebnisgrößen wie z.B. die totale Bevölkerungszahl. Mit dem Abstraktionsniveau kann zudem bestimmt werden, ob die Ergebnisgröße aggregiert oder aufgeschlüsselt nach verschiedenen

---

<sup>2</sup>In der Praxis kann es aber genau genommen nur diskrete Modelle geben, da die Berechnungsschritte immer in (wie klein auch immer) vorgegebenen Zeitintervallen ausgeführt werden.

<sup>3</sup>Ein Ergebnisverlauf einer Ergebnisgröße enthält für jedes Projektionsintervall genau einen Prognosewert.

<sup>4</sup>Es gibt noch eine Vielzahl weiterer zu unterscheidender Modellmerkmale, die jedoch für den weiteren Verlauf der Arbeit unwesentlich sind.

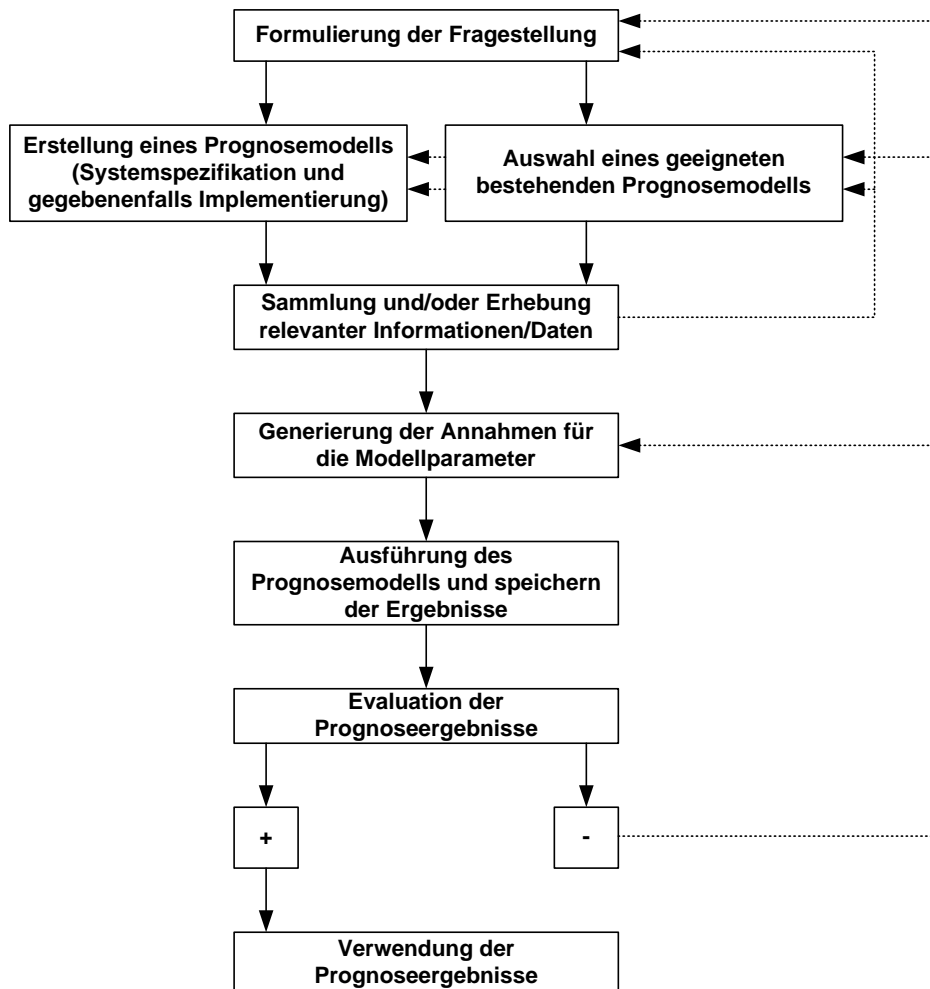


Abbildung 2.2.: Genereller Ablauf eines Prognoseprozesses von der Formulierung der Fragestellung bis zur Berechnung und Verwendung der Prognoseergebnisse (durchgezogene Pfeile); ein Fehlschlagen in einem Arbeitsschritt kann das Ansetzen an einem früheren Arbeitsschritt erfordern (gestrichelte Pfeile).

Merkmale wie z. B. dem Alter oder dem Geschlecht prognostiziert werden soll. Je niedriger das Abstraktionsniveau gewählt wird, desto höher ist der Grad der Disaggregation.

Die Ergebnisgrößen, deren Abstraktionsniveau wie auch die Länge des Projektionshorizonts entscheiden maßgeblich die Wahl des zu verwendenden Prognosemodells. Steht kein geeignetes Prognosemodell für die Umsetzung der Zielstellung zur Verfügung, muss ein neues Prognosemodell erstellt werden. Dafür müssen die notwendigen Modellparameter und

deren funktionale Zusammenhänge zur Bestimmung der entsprechenden Ergebnisgrößen bei der so genannten Systemspezifikation als Modellformalismus festgelegt werden. Mit zunehmender Komplexität des Modellformalismus und der zu treffenden Annahmen für die Modellparameter kann es außerdem erforderlich sein, einen Simulator zur Ausführung des Prognosemodells zu implementieren.

Um das ausgewählte bzw. erstellte Prognosemodell für die Bearbeitung der Zielstellung nutzen zu können, müssen alle relevanten und verfügbaren Informationen für alle Modellparameter möglichst vollständig zusammengetragen werden. Informationsquellen können z. B. in der Vergangenheit beobachtete Werte, Befragungen oder auch erklärende Theorien sein. Dabei kann es passieren, dass nicht genügend Daten zur Verfügung stehen, sodass entweder fehlende Daten extra erhoben werden oder das Prognosemodell überarbeitet oder/und die Fragestellung neu formuliert werden müssen.

Wenn alle relevanten Informationen für die Modellparameter verfügbar sind, werden die Annahmen für die zukünftige Entwicklung dieser Modellparameter mit geeigneten Methoden generiert (vgl. Kapitel 5, 6 und 7). Grundsätzlich können statistische von wertenden (und kombinierten) Methoden unterschieden werden [307]. Statistische Methoden nutzen vornehmlich historische Zeitreihenwerte, wohingegen wertende Methoden auch Informationen aus Befragungen, z. B. über die so genannte *Expertenmeinung*, bei der Generierung der Annahmen berücksichtigen. Kombinierte Methoden nutzen sowohl historische Zeitreihenwerte als auch Expertenmeinung.

Liegen alle Annahmen als Eingabewerte für jeden Modellparameter bereit, können diese in das Prognosemodell zur Ausführung eingegeben werden. Nach der Berechnung der Prognoseergebnisse können diese (in einer externen Datenbank) zur weiteren Analyse gespeichert werden.

Bevor die Ergebnisse einer Bevölkerungsprognose zu weiteren Zwecken verwendet werden, sollten sie mit entsprechenden Gütekriterien und Fehlermaßen überprüft werden (vgl. Kapitel 8). So kann z. B. *unmittelbar nach* einer Bevölkerungsprognose überprüft werden, inwiefern die Prognoseergebnisse tatsächlich die eingangs formulierte Fragestellung beantworten und inwiefern sie mit Ergebnissen vergleichbarer Bevölkerungsprognosen übereinstimmen. Im Gegensatz dazu kann die Genauigkeit der Prognoseergebnisse erst *einige Zeit nach* einer

Bevölkerungsprognose überprüft werden, indem die Abweichungen zwischen den prognostizierten und tatsächlich eingetroffenen Werten quantifiziert werden.

### 3. Entwicklung der Bevölkerungsprognose von den ersten Anfängen bis zur Gegenwart

Das Interesse an der zukünftigen Entwicklung einer Bevölkerung besteht seit einem relativ langen Zeitraum, in dem sich die Bedeutung und Methodik der Bevölkerungsprognosen verändert haben. Sind Bevölkerungsprognosen anfangs häufig zur Untermauerung aufgestellter Bevölkerungstheorien im Zusammenhang mit der potentiellen Tragfähigkeit der Erde — wie z. B. von Süßmilch [274], Malthus und Pearl und Reed [211, 230] ab dem 18. Jahrhundert — diskutiert worden [62, 63], dienen sie heute häufig als Basis wichtiger Entscheidungen in diversen gesellschaftlichen Bereichen wie z. B. der Raumplanung (vgl. Kapitel 1). Um diese Anforderungen erfüllen zu können, sind anstelle der vormaligen Bevölkerungswachstumsgesetze fortgeschrittene Methoden gerückt, die u. a. immer mehr Erklärungsfaktoren bei der Prognose der zukünftigen Bevölkerungsentwicklung verwenden. Die Wahl der zu berücksichtigenden demographischen und nicht-demographischen Erklärungsfaktoren stellt dabei ein Kriterium zur Unterscheidung bestehender Bevölkerungsprognosemethoden dar (vgl. Abschnitt 3.2 ab Seite 31). Aus der Erkenntnis, dass die Bevölkerungsentwicklung ein Resultat aus dem komplexen Zusammenspiel diverser Faktoren ist, hat sich mit der Zeit ein Bewusstsein für die Unsicherheit von Bevölkerungsprognosen entwickelt. Es sind einige Versuche unternommen worden, diese Prognoseunsicherheit mit unterschiedlichen Methoden einzufangen. Allerdings stellt die umfassende Quantifizierung der Unsicherheit einer Bevölkerungsprognose bis heute eine Herausforderung dar.

In diesem Kapitel werden Stationen der geschichtlichen Entwicklung von Bevölkerungsprognosen beleuchtet. Dabei soll der Fokus vorrangig auf der Herausbildung des Bewusstseins für die Unsicherheit von Bevölkerungsprognosen liegen.

### 3.1. Bevölkerungsprognosen in der Vergangenheit

Die Bevölkerungszahl ist bereits vor 5.000 bis 6.000 Jahren in verschiedenen Gesellschaften erhoben worden. So beispielsweise in Babylon (vor etwa 5.800 Jahren), in China (vor etwa 5.000 Jahren) oder auch in Ägypten (vor etwa 4.600 Jahren). In Rom sind die ersten regelmäßig durchgeführten Zensen in Europa dokumentiert. Nach dem Untergang des Römischen Reiches sind im Mittelalter nur noch vereinzelt Bevölkerungserhebungen wie z. B. das Güterverzeichnis von Karl dem Großen oder das *Doomsday Book* von Wilhelm dem Eroberer bekannt. Diese frühen Volkszählungen sind jedoch nicht vorrangig zur Erfassung der Bevölkerungszahl und ihrer Struktur, sondern vielmehr zur Ermittlung der Höhe der Steuern und Abgaben für jeden einzelnen Bürger wie auch zur Erfassung der Wehrpflichtigen durchgeführt worden<sup>1</sup> [63, 233].

#### 3.1.1. Erste Ansätze zur Bevölkerungsprognose

Im Laufe des 17. und 18. Jahrhunderts haben sich Wissenschaftler in Europa erstmals intensiv mit der Bevölkerungsentwicklung beschäftigt. Dabei haben sie versucht, anhand der vergangenen Bevölkerungsentwicklung Gesetzmäßigkeiten zu erkennen und diese in einem Wachstumsgesetz zu formulieren. Häufig sind damals arithmetische und geometrische Wachstumsmodelle aufgestellt worden, die zuerst graphisch und später verstärkt auch mit einer mathematischen Formel ausgeführt worden sind [230, 306]. Einerseits sollte mit den Wachstumsmodellen ein biblischer Beweis erbracht werden, der bestätigt, dass sich die Menschheit in relativ kurzer Zeit ausgehend von nur einem Paar bis zur damaligen Bevölkerungsgröße entwickeln kann. Andererseits sollten existenzielle Fragestellungen beispielsweise zum Zusammenhang zwischen der Bevölkerungsentwicklung und der Tragfähigkeit der Erde beantwortet werden können [216].

#### Süßmilch — *Göttliche Ordnung*

Süßmilch [274] hat sich 1741 mit der Bevölkerungszahl im Zusammenhang mit den Lebensbedingungen und der Tragfähigkeit der Erde befasst. Dabei ist er von einer natürlichen Selbstregulation der demographischen und ökonomischen Einflussfaktoren ausgegangen.

---

<sup>1</sup>Auch in der Gegenwart werden Zensen zur Erfassung der Bevölkerungszahl- und -struktur (in den Bereichen Bildung, Beruf und Wohnen) durchgeführt, um beispielsweise Ausgaben des Bundeshaushalts vorausschauend planen zu können; der nächste Zensus wird in Deutschland im Rahmen des *2010 World Population and Housing Census Programme* der UN im Mai 2011 durchgeführt werden [www.zensus2011.de].

Auf der Basis eigener und von anderen Wissenschaftlern durchgeführten empirischen und theoretischen Analysen ist Süßmilch zu dem Schluss gekommen, dass auf der Erde noch zehn- bis sechzehn mal mehr Menschen (besser) leben können als damals zu seiner Zeit als die Bevölkerungszahl in etwa 800 Millionen Menschen umfasst hat [25, 26, 27, 63].

1761 erscheint in der zweiten Auflage von Süßmilchs Werk der *Göttlichen Ordnung* eine von Süßmilch beauftragte Berechnung zur Bevölkerungsentwicklung ausgehend von nur einem Paar von Euler. Dabei geht Euler von konstanten, relativ unrealistischen Annahmen zur Fertilität und Mortalität aus. So soll jedes gleichaltrige Paar in den Altern 22, 24 und 26 ein ungleichgeschlechtliches Zwillingspaar bekommen. Darüber hinaus sollen alle Personen im Alter 40 sterben. Resultat ist ein geometrisches Bevölkerungswachstum, mit dem bewiesen werden kann, dass sich innerhalb eines relativ kurzen Zeitraums eine relativ große Bevölkerung von nur einem Paar aus entwickeln kann. Allerdings ist die absolute Bevölkerungsentwicklung nach wenigen hundert Jahren derart groß, dass eine zukünftige Bevölkerungszunahme nach diesem Wachstumsgesetz irrational ist [102, 216].

#### **Verdopplungsperioden**

Ein anderer Ansatz hat versucht, die Bevölkerungsentwicklung mit der Vorgabe von Verdopplungsperioden nachzubilden. Verdopplungsperioden geben die Anzahl der Jahre vor, in denen sich die Bevölkerung vom Bestand aus gesehen verdoppelt. Vertreter dieses Ansatzes, wie z. B. Sir William Petty [214], haben herausgefunden, dass sich die Verdopplungsperioden zunehmend über einen längeren Zeitraum erstrecken, sodass von einem verlangsamten Bevölkerungswachstum in der Zukunft ausgegangen werden kann [102].

#### **Malthus — Geometrisches Bevölkerungswachstum**

Malthus [180] hat eine Bevölkerungstheorie aufgestellt, in der er behauptet, dass eine ungehemmte Bevölkerung geometrisch und die lebenswichtigen Ressourcen wie die Nahrungsmittel nur arithmetisch wachsen. Demzufolge müssen das Wachstum der Bevölkerung und das der lebenswichtigen Ressourcen eine Balance finden, deren Niveau sich sowohl nach dem Entwicklungsstand der entsprechenden Technologie und Wissenschaft als auch nach dem jeweiligen Lebensstandard der Menschen bemisst. Im Gegensatz zu Süßmilch ist Malthus davon ausgegangen, dass die damalige Bevölkerungszahl der Erde schon einer Überbevölkerung gleichkomme und jeglicher weiterer Bevölkerungszuwachs zu einer Verelendung führen müsse. Diese Feststellungen haben sich im Nachhinein als falsch erwiesen [26, 27, 63, 102].



Entgegen der Behauptung Malthus, dass ein gegebener Entwicklungsstand der Nahrungsmittelproduktion das Bevölkerungswachstum bedingen und sogar stoppen kann, hat Boserup [34] 1981 bei ihren Überlegungen über das Zusammenspiel zwischen dem Wachstum der Bevölkerung und seinen lebenswichtigen Ressourcen eine weitere Determinante, und zwar die des Technologiefortschritts, mit einbezogen. So ist Boserup der festen Überzeugung, dass der durch ein überproportionales Bevölkerungswachstum entstehende Mangel an lebenswichtigen Ressourcen mit einem Technologiefortschritt ausgeglichen werde, da insbesondere existenzielle Notlagen technologische Erfindungen und Weiterentwicklungen anregen. Folglich bedingt das Bevölkerungswachstum die Nahrungsmittelproduktion und damit die Wirtschaftsentwicklung und nicht umgekehrt. Mit dieser Auffassung vertritt Boserup ein nahezu zeitlich uneingeschränktes Bevölkerungswachstum, wobei sie die Auswirkungen auf die Umwelt völlig außen vor lässt [63]. Auch Simon [253] ist der Auffassung, dass der durch ein Bevölkerungswachstum zunehmende Druck auf die verfügbaren lebenswichtigen Ressourcen (wie z. B. Nahrung und Energie) mit der Erfindung neuartiger Ideen nicht nur ausgeglichen, sondern gegenüber dem ursprünglichen Niveau sogar reduziert werden kann. Eine große Anzahl (ausgebildeter) Personen erweitere dabei das Innovationspotential, um die Ressourcenknappheit noch effektiver bekämpfen zu können.

#### **Pritchett — Polynomiales Bevölkerungswachstumsmodell**

Pritchett [222] hat als einer der ersten Wissenschaftler, die sich mit Bevölkerungsprognosen beschäftigen, die Irrationalität der bis zu diesem Zeitpunkt verwendeten meist arithmetischen und geometrischen Wachstumsmodelle erkannt. Deren Irrationalität begründet sich auf der Annahme unendlich anhaltenden und z. T. sogar unendlich zunehmenden Bevölkerungswachstums. Um diese Irrationalität aufzuheben, hat Pritchett die Verwendung eines polynomialen Bevölkerungswachstumsmodells dritten Grades vorgeschlagen, mit dem auch ein mit der Zeit abnehmendes Wachstum modelliert werden kann. Die praktische Anwendung des polynomialen Modells dritten Grades nach Pritchett für die US-Bevölkerung mit der Basisperiode von 1790 bis 1880 zeigt zwar eine relativ hohe Anpassungstreue an die Basisdaten, jedoch sind die mit diesem geschätzten Modell prognostizierten Daten schon in den ersten Projektionsintervallen immer noch deutlich zu hoch [75, 211, 216, 230].

### **Verhulst — Logistisches Bevölkerungswachstum**

Verhulst [302] und Pearl und Reed [211, 230] gehen (ähnlich wie Fourier und Quételet) davon aus, dass eine ungehemmte Bevölkerung dem geometrischen Wachstumsgesetz folgt. Jedoch kann eine Bevölkerung in einem begrenzten Gebiet (auf lange Sicht) nicht unendlich wachsen, da sie die bereits erreichte Bevölkerungsdichte wie auch die knapper werdenden Ressourcen zum Überleben immer stärker daran hindern bzw. darin einschränken. Folglich müsse einer Phase zunehmenden Bevölkerungswachstums ein Umkehrpunkt folgen, nach dem das Bevölkerungswachstum langsam bis zu einem Nullpunkt abnimmt. Ist das Bevölkerungswachstum bei *null* angelangt, hat die Bevölkerung eine obere Grenze, die der Tragfähigkeit des zu betrachtenden Gebietes entspricht, erreicht. Diese Theorie formuliert zuerst Verhulst mathematisch in Form der berühmten S-Kurve, der logistischen Wachstumskurve. Der entscheidende Unterschied zwischen den bis hierhin verwendeten mathematischen Wachstumsgesetzen und der Theorie des logistischen Bevölkerungswachstums besteht also darin, dass bei einem logistischen Wachstum sowohl eine vorher festzulegende obere Sättigungsgrenze als auch ein verlangsamtes Wachstum einer Bevölkerung modelliert werden kann. Darüber hinaus hat das logistische Wachstumsmodell eine theoretische Basis, mit der die Bevölkerungsentwicklung über einen sehr langen Zeitraum nachgebildet und prognostiziert werden kann. Im Gegensatz dazu sind die einfachen arithmetischen und geometrischen Wachstumsmodelle eher für Kurzzeitprognosen geeignet [211]. Die beiden Biologen Pearl und Reed [211, 230] haben ein logistisches Wachstumsmodell für die Prognose der US-Bevölkerung verwendet. Dabei haben sie konstatiert, dass nur unter einer Stabilitätsannahme der Wachstumsbedingungen die resultierende Bevölkerungszahl realistisch sei. Bei instabilen Wachstumsbedingungen müsse folglich jedes logistische Bevölkerungsprognosemodell aktualisiert werden, indem es an die veränderten Bedingungen angepasst wird. Demzufolge ist die Bestimmung einer oberen Sättigungsgrenze wie auch einer geeigneten Wachstumsrate für das logistische Modell durch die sich ständig verändernden Wachstumsbedingungen gerade auch in einer menschlichen Bevölkerung schwierig. Zumal hier nicht nur äußere Faktoren aus der Umwelt, sondern auch (von einem Prognostiker schwer einschätzbare) individuelle Faktoren wie beispielsweise das Geburtenverhalten zu beachten sind [75, 71, 103, 139].

### **Zusammenfassung**

Die ersten Ansätze zur Bevölkerungsprognose weisen einige Gemeinsamkeiten auf. So unterstellen sie, dass die Bevölkerungsentwicklung einem mathematischen Wachstumsgesetz folge

und deshalb mit Sicherheit vorhersagbar sei. Darüber hinaus prognostizieren diese ersten Ansätze vereinfachend die totale Bevölkerungszahl in Abhängigkeit von (hauptsächlich) ökologischen und ökonomischen Faktoren (wie z. B. der Tragfähigkeit der Erde oder auch der potentiellen Nahrungsmittelproduktion), sodass einerseits für alle Personen einer Bevölkerung dasselbe demographische Verhalten angenommen wird und andererseits keine demographischen Erklärungsfaktoren (aus den Bereichen der Fertilität, Mortalität und Migration) berücksichtigt werden. Folglich wird eine Bevölkerung auch nicht nach Alter und Geschlecht unterschieden.

#### 3.1.2. Erste moderne Ansätze zur Bevölkerungsprognose

Ab der Mitte des 19. Jahrhunderts hat die regelmäßige Sammlung, Aufbereitung und Analyse statistischer Daten zur Demographie eingesetzt und ein Umdenken in der Theorie und Methodik zur Bevölkerungsprognose bewirkt. Die zusätzliche Berücksichtigung quantitativer demographischer Entwicklungen hat nicht nur die bei einer Bevölkerungsprognose zu nutzenden Informationsquellen erweitert, sondern auch den Ablauf eines Prognoseprozesses neu strukturiert [71]. Anstatt der Suche nach einem mathematischen Wachstumsgesetz für die totale Bevölkerung werden nun bei der Erstellung eines Prognosemodells demographische Erklärungsfaktoren als Modellparameter integriert und anschließend separate Annahmen zur Fertilität und Mortalität getroffen (vgl. Abschnitt 2.2 ab Seite 18).

#### **Cannan — Entwicklung der ersten modernen demographischen Bevölkerungsprognose**

Cannan [46] ist der Begründer der modernen demographischen Bevölkerungsprognostik, da er 1895 erstmalig bei einer nationalen Bevölkerungsprognose für England und Wales die Altersstruktur der Bevölkerung wie auch die Entwicklung einzelner demographischer Komponenten, insbesondere die der Fertilität und die der Mortalität, berücksichtigt hat. Im Gegensatz zu vielen seiner Kollegen hat sich Cannan ausdrücklich von der Vorgehensweise der ersten Ansätze zur Bevölkerungsprognose (vgl. Abschnitt 3.1.1 ab Seite 23) distanziert.

Cannans neuartiger Ansatz unterscheidet Geburtskohorten, um die Bevölkerung nach Altersgruppen mit einer Breite von jeweils 10 Jahren zu prognostizieren. Für jede dieser Altersgruppen werden vom Startjahr (1891) bis zum Endjahr der Prognose (1991) Annahmen zu demographischen Ereignissen (wie beispielsweise zu Sterbefällen, Nettowanderung und Geburten) generiert. Grundsätzlich erkennt Cannan, dass jede Bevölkerungsprognose auf

Überlegungen zur zukünftigen Entwicklung wesentlicher demographischer Wachstumsfaktoren (wie Fertilität und Immigration) und Schrumpfungsfaktoren (wie Mortalität und Emigration) basieren sollte. Demzufolge formuliert er sowohl Annahmen für die Mortalität und die Nettowanderung (*rate of mortality and loss by migration*) als auch für die Fertilität. Die Annahmen für die Schrumpfungsfaktoren umfassen altersspezifische Überlebens- bzw. Sterbeanteile und altersspezifische Bevölkerungsanteile der Nettowanderung. Dabei geht Cannan generell davon aus, dass sowohl die Überlebensanteile als auch der Anteil der Nettowanderung über den Projektionshorizont in jeder Altersgruppe konstant bei den beobachteten Werten aus der Periode zwischen dem Zensus von 1881 und dem Zensus von 1891 bleibt. Diese Annahme begründet er mit der bestehenden Unsicherheit von Bevölkerungsprognosen, wodurch ein negatives und ein positives Bevölkerungswachstum gleich wahrscheinlich seien. Um diese getroffenen Annahmen empirisch zu begründen, nutzt Cannan außerdem historische Zeitreihen, die einen sinkenden Trend bei der Mortalität und einen ausgleichenden wachsenden Trend bei der Nettowanderung zeigen. Daraus resultiert die Annahme einer (annähernd) konstanten Schrumpfrate durch die Mortalität und die Nettowanderung. Für die (nach Cannan spekulativeren) Annahmen der Wachstumsfaktoren analysiert Cannan relativ lange Zeitreihen (1853 bis 1891) zu Parametern der Fertilität. Dabei setzt er die quantitative Entwicklung der Personen im fertilen Alter zwischen 20 und 40 zu der absoluten Zahl der Geburten in Beziehung. Cannan stellt ab 1876 zwar eine konstante Geburtenzahl aber auch eine sinkende Geburtenrate<sup>2</sup> in der Basisperiode fest. Folglich muss die Personenzahl im fruchtbaren Alter zugenommen haben. Deshalb geht Cannan in der Zukunft von einer Veränderung der Altersstruktur durch eine einsetzende demographische Alterung aus. Außerdem nimmt Cannan an, dass die Geburtenrate durch den gesteigerten sozialen und wirtschaftlichen Druck auf die arbeitende Bevölkerung weiter sinke. Die Annahmen der sinkenden Geburtenrate und der gleichzeitig wachsenden Überlebensanteile führen zur Annahme einer beinahe konstanten zukünftigen Geburtenzahl, die in ihrer Höhe dem Wert vom Startjahr der Prognose (1891) entspricht. Nach der Generierung der einzelnen Annahmen für die demographischen Komponenten erfolgt die Berechnung der Bevölkerungszahl und -struktur über jedes Projektionsintervall bis zum Endjahr der Prognose.

Der Vergleich der von Cannan prognostizierten mit den tatsächlich eingetroffenen Bevölke-

---

<sup>2</sup>Cannan berechnet für seine Bevölkerungsprognose die Geburtenrate, indem er die Anzahl der Geburten durch die Anzahl der Männer und Frauen im Alter zwischen 20 und 40 dividiert.

rungswerten zeigt eine relativ hohe Diskrepanz: So prognostiziert Cannan für England und Wales eine Bevölkerungszahl von 37.376.000 im Jahr 1991, wohingegen tatsächlich 50.648.190<sup>3</sup> Personen zu diesem Zeitpunkt erfasst worden sind. Allerdings besteht das grundlegende Ziel Cannans nicht in der exakten, sondern in der wahrscheinlichen Prognose der Bevölkerungszahl, wobei er insbesondere über diesen langen Projektionshorizont (anhand der potentiellen Entwicklung der Fertilität, Mortalität und Migration) ein wahrscheinliches Absinken des Bevölkerungswachstums bis zu einem Stillstand bzw. einer Umkehr aufzeigen wollte.

Mit seiner Arbeit hat Cannan den Grundstein für moderne demographische Bevölkerungsprognosen gelegt. Anstelle der Prognose der totalen Bevölkerung mit einem allgemeingültigen mathematischen Wachstumsgesetz, prognostiziert er die Bevölkerung nach Alter unter Berücksichtigung demographischer Einflussfaktoren aus den Bereichen der Fertilität, Mortalität und Migration. Für jeden dieser Einflussfaktoren kann Cannan altersspezifische Annahmen treffen, sodass der Informationsgehalt und damit auch die Genauigkeit der resultierenden Bevölkerungsprognose gegenüber den ersten Ansätzen gesteigert werden kann. Darüber hinaus ist sich Cannan über die in Bevölkerungsprognosen enthaltene Unsicherheit bewusst; so fließt z. B. die Prognoseunsicherheit teilweise in die Argumentation zur Findung geeigneter Annahmen für einzelne Modellparameter ein [46, 71].

#### **Fahlbeck — Moderner Ansatz zur Generierung von Fertilitätsannahmen**

Fahlbeck [85] hat sich zu Beginn des 20. Jahrhunderts im Gegensatz zu Cannan nicht mit der Entwicklung eines neuartigen Bevölkerungsprognosemodells, sondern mit der Generierung von Annahmen, vornehmlich für die Fertilität, beschäftigt. Entgegen der damals gängigen Überzeugung eines in der Zukunft anhaltenden Bevölkerungswachstums aus dem 19. in das 20. Jahrhundert, hat Fahlbeck einen Bevölkerungsrückgang für alle westlichen, zivilisierten Länder prognostiziert. Dabei begründet er den Fertilitätsrückgang damit, dass Frauen (insbesondere aus der Oberschicht) ihre natürliche Bereitschaft, sich zu reproduzieren, durch zunehmende Wahlfreiheit und zunehmende alternative Lebensgestaltungsmöglichkeiten, vermindern. Dadurch kann die Fertilität in der Zukunft unter das Bestandserhaltungsniveau sinken. Darüber hinaus hat Fahlbeck auch ein negatives Bevölkerungswachstum nicht ausge-

---

<sup>3</sup>Datenquelle: Human Mortality Database. University of California, Berkeley (USA), und Max Planck Institut für Demographische Forschung (Deutschland). Erhältlich unter <http://www.mortality.org> oder <http://www.humanmortality.de> (Zugriff am 29.10.2010)

schlossen, wobei er davon ausgegangen ist, dass die Geburtenrate durchaus unter die Sterberate sinken könne [71].

#### **Westergaard — Moderner Ansatz zur Generierung von Fertilitäts- und Mortalitätsannahmen**

Westergaard [303] hat sich ebenso wie Cannan und Fahlbeck zu Beginn des 20. Jahrhunderts mit der zukünftigen Bevölkerungsentwicklung auseinandergesetzt. Die Prognose der Mortalität und Fertilität und deren Auswirkungen auf die Altersstruktur haben dabei im Fokus seiner Forschungsarbeit gestanden. Westergaard ist von einem zukünftigen Rückgang sowohl in der Sterblichkeit als auch in der Fertilität ausgegangen. Als Erklärung für diese in der damaligen Zeit unübliche Annahme, die einem Bruch mit der historischen Entwicklung gleichkommt, führt er Fortschritte in der Hygiene und Medizin wie auch einen sozialen und ökonomischen Wandel an. Effekte auf die zukünftige Altersstruktur entsprechen einer demographischen Alterung, da weniger Geburten und mehr Überlebende den Anteil der Jungen verkleinern und den Anteil der Älteren vergrößern [71].

#### **Zusammenfassung**

Die ersten modernen Ansätze zur Bevölkerungsprognose nutzen sowohl bevölkerungsdynamische als auch bevölkerungstheoretische Analysen. Sie berücksichtigen neben bevölkerungsdynamischen Effekten der Fertilität, Mortalität und z. T. der Migration auch deren Effekte auf die zukünftige Alters- und Geschlechtsstruktur. Als Informationsquelle dienen ihnen historische Zeitreihen entsprechender demographischer Parameter wie auch Überlegungen und Theorien zu den Auswirkungen sozialer und ökonomischer Faktoren auf die zukünftige Bevölkerungsentwicklung. Der Einsatz bevölkerungstheoretischer Überlegungen nimmt mit der Länge des Projektionshorizonts zu. Darüber hinaus haben Vertreter dieser ersten modernen Ansätze zur Bevölkerungsprognose erstmals erkannt, dass die zukünftige Bevölkerungsentwicklung von diversen Faktoren abhängig und folglich nur mit Unsicherheit zu prognostizieren ist. Deshalb bezeichnet z. B. Westergaard [303] seine Prognose als *Horoskop* oder Cannan ([46], S. 513-514) betont, dass er nicht seinen Ruf als *Prophet* aufs Spiel setzen möchte, nur weil die tatsächliche von der von ihm aufgezeigten Bevölkerungsentwicklung in der Zukunft leicht abweiche. Demnach beabsichtigt Cannan nicht die zukünftige Bevölkerungsentwicklung exakt zu prognostizieren, sondern die grobe bzw. wahrscheinliche Entwicklung dieser aufzuzeigen [71].

### 3.1.3. Übergang von den Bevölkerungsprognosen aus der Vergangenheit in die Gegenwart

Mit dem Aufkommen der ersten modernen, demographischen Ansätze zur Bevölkerungsprognose ist erstmals auch ein Gespür für die Unsicherheit von Bevölkerungsprognosen entwickelt worden. Allerdings haben sie sich nicht widerstandslos gegen die damals auch noch weit verbreiteten ersten, mathematischen Ansätze zur Bevölkerungsprognose durchsetzen können. So ist vorrangig das logistische Wachstumsmodell (z. B. von Pearl und Reed [211, 230]) noch über einen längeren Zeitraum relativ stark vertreten worden. Jedoch hat die Idee, die zukünftige Bevölkerungsentwicklung nach demographischen Ereignissen wie den Geburten, Sterbefällen und Wanderungen unter bestimmten sozialen, ökonomischen und biologischen Bedingungen zu prognostizieren, immer mehr Wissenschaftler überzeugen können, sodass sich letztendlich der moderne, demographische Ansatz zur Bevölkerungsprognose in den dreißiger Jahren des 20. Jahrhunderts international durchsetzen konnte. Als Folge ist er dann auch in vielen Ländern zur Erstellung nationaler und regionaler Bevölkerungsprognosen verwendet worden [71]. Die ungleichgewichtige Berücksichtigung der Fertilität, Mortalität und Migration stellt aber bei den ersten praktischen Anwendungen des modernen, demographischen Ansatzes ein Problem dar. Während die Fertilität im Zentrum der Aufmerksamkeit gestanden hat, sind die Mortalität und stärker noch die Migration vernachlässigt worden. Erklärungen bieten einerseits die damals (überraschend) schrumpfende Fertilitätsentwicklung in etlichen westeuropäischen Ländern und andererseits die relativ stabile Mortalitätsentwicklung wie auch die geringe Datenbasis für die Migration derzeit [30, 72, 71, 103, 119, 270].<sup>4</sup>

## 3.2. Bevölkerungsprognosen in der Gegenwart

### 3.2.1. Kohorten-Komponenten Methode

Die Kohorten-Komponenten Methode ist erstmalig von Cannan [46] eingeführt und ungefähr 30 Jahre später unabhängig von Bowley [35] und Whelpton [304, 305] noch einmal entwickelt worden. Die Bevölkerungsprognosen von Whelpton nach dem *analytischen Ansatz* haben

---

<sup>4</sup>Diese Sicht- und Vorgehensweise hängt mit der damaligen Einschätzung der Eignung der einzelnen demographischen Komponenten für eine Vorausschätzung zusammen. So ist gerade die Migration als ungeeignet für Prognosen erachtet worden, da hier — wie auch in geringerem Maß bei der Fertilität und der Mortalität — die *Spontanität des menschlichen Willens* besonders entwicklungsbestimmend sei (Deneffe 1938 [72], S. 19-20).

eine sehr breite Akzeptanz und eine hohe Popularität gefunden<sup>5</sup> [75, 256].

Bei der Kohorten-Komponenten Methode handelt es sich um die am meisten verbreitete Methode zur Generierung von Bevölkerungsprognosen. Ihre Besonderheit besteht darin, möglichst homogene Subpopulationen, die neben dem Alter nach verschiedenen Merkmalen wie z.B. dem Geschlecht, der Region oder auch der Ethnizität variabel gebildet werden können, separat über spezifische Annahmen zur Fertilität, Mortalität und Migration zu prognostizieren. Neben der zukünftigen Bevölkerungszahl können somit auch Aussagen zur zukünftigen Bevölkerungsstruktur getroffen werden. Ferner können die Beiträge der einzelnen demographischen Komponenten zum resultierenden Bevölkerungswachstum ermittelt werden.

Seit der Einführung der Kohorten-Komponenten Methode sind zahlreiche Weiterentwicklungen entstanden, die sich jeweils an die gegebene Fragestellung und Rahmenbedingungen der zu erstellenden Prognose angepasst haben. So ist beispielsweise der Detailgrad der Modellierung von Migration, also z.B. die Betrachtung von Brutto- oder Nettomigration, ein variierender Aspekt unter den verschiedenen bestehenden Varianten der Kohorten-Komponenten Methode. Das grundlegende System der jahrgangsweisen Fortschreibung für einzelne Subpopulationen ist jedoch gleich geblieben.<sup>6</sup>

#### **Klassische Kohorten-Komponenten Methode**

Für eine Bevölkerungsprognose mit der Kohorten-Komponenten Methode muss in einem ersten Schritt festgelegt werden, ob die zu prognostizierende Bevölkerung nach Einzelalter oder aggregierten Altersgruppen prognostiziert werden soll. Diese Entscheidung bedingt auch die Länge der Projektionsintervalle, die gemeinhin mit der Länge der Altersklassen bzw. mit einem Vielfachen davon übereinstimmen.

Anschließend wird in Abhängigkeit von der Fragestellung die zu prognostizierende Bevölkerung aus dem Startjahr der Prognose nach weiteren Merkmalen wie dem Geschlecht

---

<sup>5</sup>In ihren Grundzügen entspricht die Kohorten-Komponenten Methode der *Biologischen Methode* vom Statistischen Reichsamt [215]. Ausführungen zur *Biologischen Methode* sind u. a. bei Deneffe [72], Hage [103] oder auch bei einer Veröffentlichung des Statistischen Reichsamts [270] zu finden.

<sup>6</sup>Im Prinzip wird die Bevölkerung auf der Basis der Bilanzgleichung [221] fortgeschrieben, nach der sich der aktuelle Bevölkerungsbestand aus dem zuletzt beobachteten Bevölkerungsbestand ergibt, dem die seit diesem Zeitpunkt stattfindenden Geburten und Zuwanderer hinzugerechnet und von dem die seit diesem Zeitpunkt stattfindenden Sterbefälle und Fortzüge abgezogen werden.



und der Region in möglichst homogene Subpopulationen eingeteilt.

Für jeden Jahrgang dieser Subpopulationen werden gesondert Annahmen für Geburten, Sterbefälle, Immi- und Emigranten über den gesamten Projektionshorizont getroffen [305]. Spezielle Vorgaben zur Generierung dieser Annahmen gibt es nicht. Die Wahl einer oder mehrerer statistischer oder/und wertender Methoden zur Erstellung der Annahmen (vgl. Kapitel 5, 6 und 7) ist demnach ein Charakteristikum zur Unterscheidung moderner Bevölkerungsprognoseansätze, die auf der Kohorten-Komponenten Methode basieren.

Tatsächlich treten demographische Ereignisse kontinuierlich auf. In einem diskreten Bevölkerungsprognosemodell müssen jedoch genaue Zeitpunkte festgelegt werden, zu denen Geburten, Sterbefälle und Wanderungen auftreten (vgl. Abschnitt 2.1 ab Seite 16). Dies stellt eine Vereinfachung der Realität dar, die bezüglich der zu bestimmenden Reihenfolge der demographischen Ereignisse in einem Projektionsintervall wesentliche Auswirkungen auf die Prognoseergebnisse haben kann. Jedoch werden weder die Reihenfolge der demographischen Ereignisse noch die genauen Modellparameter (und die Methoden zur Generierung ihrer Annahmen) von der Kohorten-Komponenten Methode fest vorgegeben, sodass diese bei jeder Anwendung variieren können. Nach Smith et al. [256] besteht ein möglicher Ablauf einer Bevölkerungsprognose nach der Kohorten-Komponenten Methode darin, zuerst die Mortalität mit Überlebensraten, dann die Migration mit Migrationsraten und abschließend die Fertilität mit Fertilitätsraten in jedem Projektionsintervall auf die Startbevölkerung bzw. auf die Endbevölkerung des vorhergehenden Projektionsintervalls nacheinander anzuwenden (bis das Zieljahr der Prognose erreicht ist). Demnach werden in einem Projektionsintervall zunächst von der jeweiligen Ausgangsbevölkerung die Überlebenden für jedes Alter (bis auf die jüngste Altersklasse bzw. -gruppe) in jeder Subpopulation berechnet. Anschließend werden die zu prognostizierenden Immi- und Emigranten hinzugerechnet bzw. abgezogen. Abschließend werden die Neugeborenen anhand der verbleibenden Frauen im reproduktiven Alter und den entsprechenden Fertilitätsraten bestimmt. Die ermittelten Neugeborenen werden dann nach anzunehmenden Anteilen (wie z. B. dem Geschlechteranteil bei Geburt<sup>7</sup> zur Bestimmung des Geschlechts) auf die einzelnen Subpopulationen aufgeteilt und anschließend der subpopulationsspezifischen Mortalität und Migration ausgesetzt. Potentielle Variationen der Kohorten-Komponenten Methode sind beispielsweise die Verwendung zusammengesetzter Mortalitäts- und Migrationsraten oder auch eine veränderte Reihenfolge

---

<sup>7</sup>vgl. Abschnitt 9.2.5 auf Seite 175

der demographischen Ereignisse, bei der zunächst die Mortalität und die Fertilität auf die jeweilige Ausgangsbevölkerung angewendet werden und anschließend die Migration in absoluten Zahlen (netto oder auch brutto) hinzugerechnet wird.

In verschiedenen Anwendungen der Kohorten-Komponenten Methode wird die Fertilität vereinfachend allein den Frauen zugeschrieben, weshalb sie dann auch als *female-dominant models* bezeichnet werden [221]. Mit Hilfe von Annahmen zur Entwicklung des Geschlechteranteils bei Geburt können dann aber sowohl für die weibliche als auch für die männliche Subpopulation die Bestände in der untersten Altersgruppe berechnet und zugeordnet werden.

Die Prognose einzelner Subpopulationen ermöglicht u. a. die Ausgabe vielfältiger Ergebnisse. So können diverse Ergebnisgrößen in verschiedenen Abstraktionsgraden wie z. B. die Bevölkerungszahl insgesamt oder nach Geschlecht berechnet werden.

Die Vorzüge der Kohorten-Komponenten Methode liegen in ihrer Anlehnung an den natürlichen bevölkerungsdynamischen Aufbau einer Bevölkerung wie auch in ihrer generellen Anpassungsfähigkeit an verschiedene Prognosebedingungen. Die separate Betrachtung der Fertilität, Mortalität und Migration ermöglicht es, deren unterschiedliche Reaktionen auf Einflüsse relevanter gesellschaftlicher und wissenschaftlicher Bereiche (wie z. B. aus der Medizin, der Politik oder auch aus der Wirtschaft) bei der externen Generierung der Annahmen mit verschiedenen Methoden zu berücksichtigen.<sup>8</sup> Darüber hinaus kann demographische Heterogenität, die sich in unterschiedlichen Niveaus für die Fertilität, Mortalität und Migration bei Personen unterschiedlicher Merkmalskombinationen ausprägt, durch die Bildung möglichst homogener Subpopulationen (zumindest eingeschränkt) berücksichtigt werden. So bietet gerade die Unterscheidung der demographisch wichtigen Merkmale Alter und Geschlecht die Möglichkeit, die demographische Heterogenität teilweise einzufangen und in einer Bevölkerungsprognose zu einem gewissen Grad zu integrieren. Die dieser Methode eigene Flexibilität kann sich sowohl positiv als auch negativ auswirken. So ist es dem Prognostiker überlassen, ob er bei einer Prognose theoretische Kenntnisse einbaut, wie viele unterscheidende Merkmale er zur Bildung homogener Subpopulationen verwendet, welche

---

<sup>8</sup>So ist beispielsweise bei der Generierung der Annahmen für die Fertilität eine Veränderung in der Familienpolitik (wie z. B. der Beschluss des Tagesbetreuungsausbaugesetzes von 2005 [43] oder auch die Einführung des Elterngelds 2007 [44] in Deutschland) von Bedeutung, wohingegen die Entwicklung eines neuen Medikaments zur Behandlung bösartiger Neubildungen bei der Generierung der Annahmen für die Mortalität entscheidend ist.

(verschiedenen) Methoden er zur Generierung der Annahmen heranzieht, wie viele detaillierte Daten er nutzt und ob er die potentielle Vielfältigkeit der Ergebnisausgabe ausschöpft. Je nachdem, wie viele dieser Besonderheiten in eine Prognose mit der Kohorten-Komponenten Methode eingebaut werden, nimmt deren Aufwand hinsichtlich der benötigten Daten und der durchzuführenden Rechenprozeduren zu. Infolge dieser Überlegungen wird ersichtlich, dass diese grundsätzlich freie Wahl entscheidender Spezifika bei dieser Methode auch ein enormes Fachwissen beim Prognostiker voraussetzt.<sup>9</sup> Ein häufig gegen die Kohorten-Komponenten Methode vorgebrachter Kritikpunkt ist ihr ahistorischer Charakter, da sie ausschließlich Daten aus der Gegenwart für die Startbevölkerung und Daten aus der Zukunft für die Annahmen in ihre Berechnungen einschließt. Die Berücksichtigung historischer Daten, die u.a. auch die zukünftige Entwicklung der vitalen Raten mitbestimmen, kann nur indirekt über die jeweils vom Prognostiker zu verwendende Methode zur Generierung der Annahmen erfolgen. Außerdem berücksichtigt die Kohorten-Komponenten Methode direkt ausschließlich demographische Faktoren, die die zukünftige Entwicklung einer Bevölkerung beeinflussen, obwohl auch nicht demographische Faktoren diese beeinflussen können (vgl. Abschnitt 3.2.4 ab Seite 47) [63, 256, 273].

#### **Transformation der klassischen Kohorten-Komponenten Methode in ein Matrixmodell**

Die Kohorten-Komponenten Methode ist nach der Entwicklung ihrer ursprünglichen Form zur vereinfachten Darstellung und Berechnung in ein Matrixmodell umgeschrieben worden.

Die ersten *matrix population models* sind voneinander unabhängig in den 1940ern entstanden. Bernadelli [20], Lewis [166] und vor allem Leslie [162, 163, 165, 164] haben erste bedeutende Arbeiten zu Matrixmodellen eingeführt. Doch bis zu ihrer Neuentdeckung in den 1960er Jahren von u. a. Keyfitz [134, 135], Rogers [234] und Goodman [99] sind Matrixmodelle weitestgehend als Alternative zu Bevölkerungsmodellen, die auf Integral- und Differentialgleichungen basieren, vernachlässigt worden [53, 71, 137].

In einem Matrixmodell der Kohorten-Komponenten Methode wird jede der vorhandenen,

---

<sup>9</sup>Wobei Prognoseergebnisse auch trotz eines großen Fachwissens beim Prognostiker aufgrund der Prognoseunsicherheit (vgl. Kapitel 4) von der tatsächlich in der Zukunft beobachtbaren Bevölkerungsentwicklung abweichen können.

nach Altersklassen gegliederten Subpopulationen von einem Zeitpunkt  $t$  zum nächsten Zeitpunkt  $t + 1$  mit jeweils subpopulationsspezifischen Überlebenswahrscheinlichkeiten und Fertilitätsraten nach Alter (und einem entsprechenden Geschlechteranteil bei Geburt<sup>10</sup>) prognostiziert. Dazu werden die einzelnen, nach dem Alter unterteilten Subpopulationen zum Zeitpunkt  $t$  in einem Bevölkerungsvektor  $P$  und die alters- und subpopulationsspezifischen Überlebenswahrscheinlichkeiten und Fertilitätsraten (wie auch der Geschlechteranteil bei Geburt) in einer nach Leslie benannten *population projection matrix*, der Leslie-Matrix,  $L$ , zusammengetragen. Die Multiplikation der Leslie-Matrix und des Bevölkerungsvektors zum Zeitpunkt  $t$  ergeben die jeweilige Zusammensetzung einer Subpopulation nach dem Alter zum Zeitpunkt des nächst folgenden Projektionsintervalls,  $t + 1$ , in einem weiteren Bevölkerungsvektor (vgl. Formel 3.1).

$$P_{t+1} = L \cdot P_t \quad (3.1)$$

Der Aufbau der Bevölkerungsvektoren und der Leslie- oder Projektionsmatrix richtet sich nach der Anzahl der zu unterscheidenden Subpopulationen und Altersklassen. Wird eine Bevölkerung neben dem Alter nicht nach weiteren Merkmalen unterschieden, entstehen Bevölkerungsvektoren und eine quadratische Projektionsmatrix, deren Dimensionslänge der Anzahl der Altersklassen entspricht.

In der ersten Zeile enthält die Projektionsmatrix Angaben zur Bestimmung der Neugeborenen. Dies entspricht jeweils einer Kombination aus den entsprechenden altersspezifischen Fertilitätsraten,  $f_x$ , (dem jeweiligen Geschlechteranteil bei Geburt) und den entsprechenden altersspezifischen Überlebenswahrscheinlichkeiten für die Frauen im reproduktiven Alter wie auch für die Neugeborenen im ersten Lebensjahr. Des Weiteren stehen in der Subdiagonalen der Projektionsmatrix die altersspezifischen Überlebenswahrscheinlichkeiten,  ${}_1p_x$ , zur Bestimmung der Personen, die das Alter  $x$  zum Zeitpunkt  $t$  überleben und so zum Zeitpunkt  $t + 1$  in das Alter  $x + 1$  übergehen. Die restlichen Positionen der Projektionsmatrix sind mit Nullen gefüllt. Die Abbildung 3.1 zeigt den Aufbau der beiden Bevölkerungsvektoren,  $P_t$  und  $P_{t+1}$ , und der Leslie-Matrix,  $L$ , für eine menschliche Bevölkerung nach Einzelaltern bis zur höchsten, nach oben offenen Altersgruppe 100+. In der ersten Zeile der Leslie-Matrix stehen vereinfachend die Fertilitätsraten der reproduktiven Alter 15 bis 45 zur Bestimmung der Neugeborenen und in der Subdiagonalen stehen die Überlebenswahrscheinlichkeiten für alle Einzelalter zur Bestimmung der Überlebenden. Sobald sich der Bestand der 100+-Jährigen

---

<sup>10</sup>—sofern überhaupt nach dem Geschlecht unterschieden wird—

$$\begin{pmatrix} P_{0,t+1} \\ P_{1,t+1} \\ \vdots \\ P_{14,t+1} \\ P_{15,t+1} \\ \vdots \\ P_{45,t+1} \\ P_{46,t+1} \\ \vdots \\ P_{100+,t+1} \end{pmatrix} = \begin{pmatrix} 0 & 0 & \dots & 0 & f_{15} & \dots & f_{45} & 0 & \dots & 0 \\ {}_1p_0 & 0 & \dots & 0 & 0 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \ddots & \dots & \vdots & \vdots & \dots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \ddots & 0 & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & {}_1p_{14} & 0 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots & \ddots & \dots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & 0 & 0 & \ddots & 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 0 & \dots & {}_1p_{45} & 0 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots & \vdots & \dots & \vdots & \ddots & \dots & \vdots \\ 0 & 0 & \dots & 0 & 0 & \dots & 0 & 0 & \ddots & 0 \end{pmatrix} \cdot \begin{pmatrix} P_{0,t} \\ P_{1,t} \\ \vdots \\ P_{14,t} \\ P_{15,t} \\ \vdots \\ P_{45,t} \\ P_{46,t} \\ \vdots \\ P_{100+,t} \end{pmatrix}$$

Abbildung 3.1.: Kohorten-Komponenten Methode in Matrixnotation: Aufbau der Projektionsmatrix und der entsprechenden Bevölkerungsvektoren; in diesem Beispiel wird eine Bevölkerung nach Einzelalter von 0 bis 100+ für ein weiteres Projektionsintervall prognostiziert.

aber nicht nur aus Personen zusammensetzt, die vom Alter 99 neu in diese Altersgruppe eintreten, sondern auch aus Personen, die ein weiteres Projektionsintervall in dieser Altersgruppe überleben, ist zusätzlich noch eine Überlebenswahrscheinlichkeit für die Personen in dieser nach oben offenen Altersgruppe,  $p_{100+}$ , zu bestimmen; die Überlebenswahrscheinlichkeit  $p_{100+}$  ist dann an die Position der letzten Zeile und Spalte in der Projektionsmatrix einzugeben.

Wenn anstatt der Gesamtbevölkerung einzelne Subpopulationen prognostiziert werden sollen, kann die Projektionsmatrix entsprechend erweitert werden. Bei der Prognose einer Bevölkerung nach Alter und Geschlecht, bei der die Bevölkerungsvektoren in der oberen bzw. unteren Hälfte die Bestände der Frauen bzw. Männer enthalten, verdoppelt sich die Anzahl der Zeilen und Spalten in der Projektionsmatrix, die sowohl die Fertilitätsannahmen für die Frauen als auch die Mortalitätsannahmen für die Frauen und Männer enthält. Darüber hinaus können mit einer Projektionsmatrix aber auch Subpopulationen prognostiziert werden, zwischen denen Übergänge erwartet werden. So kann beispielsweise eine Bevölkerung nicht nur nach Alter und Geschlecht, sondern zudem noch nach Bildungsgrad oder Familienstand

prognostiziert werden. Dabei stellen z.B. Eheschließungen, Ehescheidungen oder auch Bildungsabschlüsse Ereignisse dar, die Übergänge zwischen den Subpopulationen auslösen.

Wenn zudem nicht eine geschlossene, sondern eine offene Bevölkerung prognostiziert werden soll, muss zusätzlich Migration berücksichtigt werden. Die Berücksichtigung von Migration stellt dabei ein wesentliches Unterscheidungsmerkmal bestehender Ansätze zur Bevölkerungsprognose dar. Die Formeln 3.2 bis 3.4 zeigen drei Varianten, Nettomigration (NI) zu unterschiedlichen Zeitpunkten in einer Prognose mit der Projektionsmatrix zu berücksichtigen; so kann die Nettomigration z. B. zu Beginn, am Ende oder jeweils zur Hälfte zu Beginn und am Ende des Projektionsintervalls (von  $t$  nach  $t + 1$ ) stattfinden. Vereinfachend werden hier (wie in vielen bestehenden Bevölkerungsprognosemodellen) dieselben Annahmen zur Fertilität und Mortalität für die Einheimischen und die Nettomigranten unterstellt.

$$P_{t+1} = L \cdot (P_t + NI_t) \quad (3.2)$$

$$P_{t+1} = L \cdot P_t + NI_t \quad (3.3)$$

$$P_{t+1} = L \cdot (P_t + \frac{NI_t}{2}) + \frac{NI_t}{2} \quad (3.4)$$

Neben der Nettomigration kann aber z.B. auch Bruttomigration mit separaten Annahmen zur Fertilität und Mortalität in einer Prognose mit der Projektionsmatrix berücksichtigt werden (vgl. Formel 3.5). Dazu werden subpopulationsspezifische Projektionsmatrizen mit separaten Fertilitäts- und Mortalitätsannahmen erstellt und mit den entsprechenden (Sub)Bevölkerungsvektoren für die Einheimischen (E), Immigranten (Im) und Emigranten (Em) multipliziert. Die Summe aller drei (Sub)Bevölkerungsvektoren ergibt dann den (Gesamt)Bevölkerungsvektor.

$$P_{t+1} = L^{Einheimische} \cdot E_t + L^{Immigranten} \cdot Im_t - L^{Emigranten} \cdot Em_t \quad (3.5)$$

Darüber hinaus hat Rogers [234] ein interregionales Prognosemodell vorgeschlagen, bei dem die Projektionsmatrix so erweitert wird, dass Wanderungsbewegungen zwischen diesen beiden Regionen berücksichtigt werden können.

Die Annahme konstanter Fertilität und Mortalität erlaubt zudem eine relativ einfache Bestimmung wesentlicher bevölkerungsdynamischer Parameter wie beispielsweise der gleich-

gewichtigen Wachstumsrate und der gleichgewichtigen Altersstruktur über die Berechnung des dominanten Eigenwerts und des dazugehörigen rechten Eigenvektors der Projektionsmatrix [53, 137, 221].

Die bisher vereinfachende Annahme konstanter Fertilität und Mortalität wird jedoch mit zunehmender Länge des Projektionshorizonts unplausibel. Während die Projektionsmatrix bei zeitinvarianten Annahmen unverändert bleibt, muss sie bei zeitvarianten Annahmen für die entsprechenden Projektionsintervalle verändert bzw. angepasst werden (vgl. Formel 3.6). Wie diese Anpassung der Projektionsmatrix erfolgt, hängt vom jeweiligen Prognoseansatz ab (vgl. Kapitel 4).

$$P_{t+n} = L_{t+n-1 \rightarrow t+n} \cdot \dots \cdot L_{t \rightarrow t+1} \cdot P_t \quad (3.6)$$

Diese verschiedenen, jeweils an bestimmte Bedingungen angepassten Varianten der Kohorten-Komponenten Methode unterscheiden sich zwar in ihren Ausführungen (wie z. B. den Berechnungsformeln oder auch dem Aufbau der Projektionsmatrix), allerdings bleibt das generelle Prinzip stets erhalten [137, 221].

#### **Funktionale Bevölkerungsprognosen auf der Basis der Kohorten-Komponenten Methode**

*Funktionale* Bevölkerungsprognosen teilen die nach der Kohorten-Komponenten Methode zu prognostizierende, alters- und geschlechtsspezifische Bevölkerung auf funktionale Zustände auf. Funktionale Zustände können beispielsweise den Erwerbsstatus,<sup>11</sup> den Haushaltstyp<sup>12</sup> oder auch den Gesundheitszustand<sup>13</sup> betreffen. Die Aufteilung der Bevölkerung auf die verschiedenen, jeweils zu betrachtenden funktionalen Zustände kann mit verschiedenen Methoden erfolgen. Generell muss nur für jede Alters- und Geschlechtsgruppe in jedem Projektionsintervall der Anteil für jeden zu betrachtenden funktionalen Zustand vorher bestimmt werden. Prävalenzraten stellen eine Möglichkeit zur Bestimmung solcher Anteilswerte dar. Die jeweils vorher zu bestimmende Verteilungsfunktion zur Aufteilung der Bevölkerung auf die jeweils zu betrachtenden Zustände kann sich zwar über den Projektionshorizont verändern. Den-

---

<sup>11</sup>z. B. erwerbstätig oder arbeitslos

<sup>12</sup>z. B. 1-Personen-Haushalt oder Mehrpersonen-Haushalt

<sup>13</sup>z. B. gesund oder krank

noch ist die hier beschriebene Vorgehensweise funktionaler Bevölkerungsprognosen statisch, da die Aufteilung auf die funktionalen Zustände erst im Anschluss an die bereits durchzuführende Bevölkerungsprognose erfolgt [28, 145, 310]. So findet z. B. keine dynamische Interaktion zwischen den einzelnen Subpopulationen nach Alter, Geschlecht und funktionalem Status statt. Dieser Nachteil kann erst bei Bevölkerungsprognosen mit der multidimensionalen Kohorten-Komponenten Methode aufgehoben werden. Allerdings sind dann auch der Modellierungs-, Daten- und Rechenaufwand deutlich erhöht. Demnach stellen funktionale Bevölkerungsprognosen eine einfache Alternative zur Prognose einer Bevölkerung nach verschiedenen Zuständen dar.

#### **Erweiterung der Kohorten-Komponenten Methode um die Multidimensionalität**

Homogene Bevölkerungen können im Vergleich zu heterogenen Bevölkerungen genauer prognostiziert werden, da bei ihnen die Annahmen für die demographischen Komponenten auf fast alle Individuen weitestgehend zutreffen. Da größere Bevölkerungen für gewöhnlich heterogen sind, können sie nach bestimmten Merkmalen wie dem Geschlecht, dem Alter, der Ethnizität oder auch der Lokalität disaggregiert werden. Diese in sich homogenen Subpopulationen können dann nach der klassischen Kohorten-Komponenten Methode prognostiziert und am Ende zusammengeschlossen werden. Die Besonderheit des *multi-state modelling*, der Erweiterung der klassischen Kohorten-Komponenten Methode um zusätzliche Dimensionen liegt nun darin, nicht nur einzelne Subpopulationen separat, sondern in einem komplexen Zusammenspiel zu prognostizieren. Neben der Fertilität, Mortalität und Migration bestimmt nun auch die Mobilität zwischen den Subpopulationen die bevölkerungsdynamische Entwicklung. Übergänge zwischen den Subpopulationen werden durch merkmalsabhängige Übergangsraten bestimmt [30, 112, 113, 147, 169, 307]. Es können grundsätzlich interne von externen Ereignissen unterschieden werden. Interne Ereignisse bezeichnen Übergänge zwischen Subpopulationen innerhalb der zu prognostizierenden Bevölkerung. Wohingegen externe Ereignisse Eintritte und Austritte über die Grenzen der zu prognostizierenden Bevölkerung bezeichnen. Die Immigration ist ein typisches Beispiel für einen externen Eintritt und die Emigration oder auch der Tod sind typische Beispiele für externe Austritte [290, 291, 292]. Rogers [235, 236] gilt als der Begründer der multidimensionalen Kohorten-Komponenten Methode, die er anfänglich für regionale Bevölkerungsprognosen konstruiert hat (vgl. Abschnitt 7.1.1 ab Seite 135). Darauf folgend sind zahlreiche weitere Anwendungen für die multidimensionale Kohorten-Komponenten Methode entstanden. So gibt es z. B. Anwendungen, die eine Bevölkerung nach dem Familienstand [290, 291, 292] oder auch nach dem Bildungsstand



[171, 172] mit der multidimensionalen Kohorten-Komponenten Methode prognostizieren.

### **Kohorten-Komponenten Methode als Basis für neue Ansätze zur Bevölkerungsprognose**

Die Kohorten-Komponenten Methode ist ab der Mitte des 20. Jahrhunderts regelmäßig von statistischen Bundesämtern verschiedener entwickelter Länder zur Erstellung nationaler Bevölkerungsprognosen genutzt worden. Jedoch haben Überprüfungen zur Genauigkeit der Prognoseergebnisse (vgl. Abschnitt 8.1 ab Seite 147) oft nur eine beschränkte Vorhersagekraft feststellen können; daran hat auch nicht der verstärkte Einsatz fortgeschrittener Methoden zur Generierung der Annahmen für die Modellparameter (wie z. B. die Prognose der Fertilität nach Kohorten oder/und der Parität (vgl. Kapitel 6)) wesentlich etwas ändern können. Ein Grund dafür ist, dass es bis zum Ende der 1970er Jahre keine signifikante Weiterentwicklung in der eigentlichen Methodik zur Bevölkerungsprognose gegeben hat [30, 125].

Dennoch sind seit der Entwicklung der Kohorten-Komponenten Methode und ihrer zahlreichen Erweiterungen bis in die aktuelle Gegenwart diverse Ansätze zur Bevölkerungsprognose entstanden. Diese unterscheiden sich in der Berücksichtigung der Prognoseunsicherheit, der Untersuchungseinheit und im Einbau exogener, nicht-demographischer Einflussfaktoren bei einer Bevölkerungsprognose. Darüber hinaus können Prognosen auch noch hinsichtlich ihrer Zielstellung unterschieden werden.

#### **3.2.2. Deterministische und probabilistische Ansätze zur Bevölkerungsprognose**

Im Hinblick auf die Berücksichtigung von Unsicherheit können deterministische von probabilistischen (oder stochastischen) Bevölkerungsprognosen unterschieden werden.

Der allgemeine Vorgang einer deterministischen Bevölkerungsprognose ist in der Abbildung 3.2 graphisch veranschaulicht. Für jeden Modellparameter des Prognosemodells muss genau ein Annahmenverlauf generiert werden, der die angenommenen Werte über den gesamten Projektionshorizont enthält. Der (in Abhängigkeit von der Komplexität des Prognosemodells eventuell einzusetzende) Simulator (zur Berechnung) des Prognosemodells nutzt dann diese eingegebenen Annahmenverläufe für alle Modellparameter, um genau einen Ergebnisverlauf für jede Ergebnisgröße zu berechnen. Ein Ergebnisverlauf enthält dabei die prognostizierten

Werte der jeweiligen Ergebnisgröße über den gesamten Projektionshorizont.

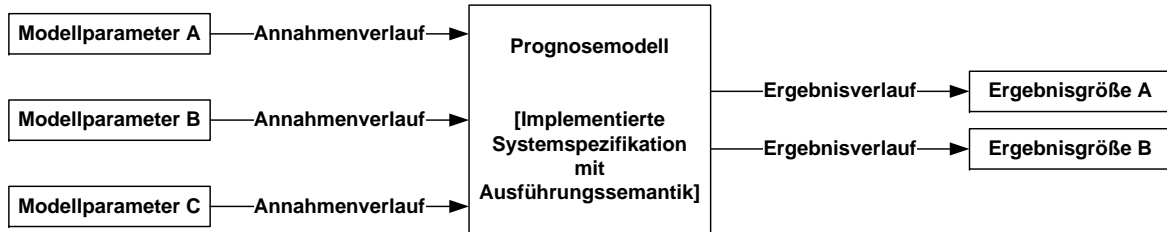


Abbildung 3.2.: Ablauf einer deterministischen Bevölkerungsprognose. Mit genau einem Annahmenverlauf pro Modellparameter wird genau ein Ergebnisverlauf für jede Ergebnisgröße prognostiziert.

Deterministische Bevölkerungsprognosen sind weit verbreitet. Jedoch können sie durch die Berücksichtigung nur eines Annahmenverlaufs pro Modellparameter das Spektrum potentieller zukünftiger Entwicklungen (für die Modellparameter wie auch für die Ergebnisgrößen) nur ungenügend einfangen. Häufig kann die potentielle zukünftige Entwicklung aber nur mit mehreren Annahmenverläufen umfassend abgedeckt werden, um die Prognoseunsicherheit angemessen einzufangen; jeder potentielle Annahmenverlauf kann dabei auf Informationen verschiedener Quellen (wie z. B. der empirischen Entwicklung, Surveys oder auch erklärender Theorien) beruhen. Anstatt sich für einen dieser potentiellen Annahmenverläufe zu entscheiden, sollten mehrere (bzw. alle) in einem Prognosemodell berücksichtigt werden können.

Eine Möglichkeit, diese Prognoseunsicherheit bzw. die potentiellen Entwicklungen der Modellparameter einzufangen besteht deshalb darin, anstatt nur eines mehrere Szenarien mit der so genannten *Szenario-Technik* zu erstellen. Jedoch stellt jedes Szenario immer noch eine deterministische Bevölkerungsprognose in sich dar, sodass die Unsicherheit zwar etwas mehr, aber noch immer ungenügend erfasst wird. Zumal den unterschiedlichen Ergebnisverläufen aus den verschiedenen Szenarien keine Eintrittswahrscheinlichkeit zugeordnet werden kann.

Nationale Statistikämter und öffentliche Institutionen greifen bei der Erstellung einer Bevölkerungsprognose häufig auf die Szenario-Technik zurück.<sup>14</sup> Dabei erstellen sie meist drei

---

<sup>14</sup>Die Szenario-Technik ist in Vergangenheit und Gegenwart z. B. vom Statistischen Bundesamt in Deutschland für die 11. und 12. koordinierte Bevölkerungsprognose [264, 266] wie auch von den Vereinten Nationen, Eurostat und der Weltbank praktiziert worden [30].

(oder mehr) alternative Entwicklungsverläufe für die zukünftige Bevölkerung. Pro Szenario kann nur ein Annahmenverlauf pro Modellparameter generiert werden, jedoch können diese Annahmenverläufe pro Modellparameter von Szenario zu Szenario variieren. Die daraus entstehenden, alternativen Ergebnisverläufe pro Ergebnisgröße können unterschiedlich interpretiert werden. So können alle alternativen Ergebnisverläufe einerseits als gleichwertig betrachtet werden oder es kann andererseits ein Ergebnisverlauf als wahrscheinlichster Verlauf angegeben werden [30, 169, 273]. Zusätzlich können die Extremverläufe aller Szenarien mit dem wahrscheinlichsten Ergebnisverlauf als Prognose mit Bereichsangabe interpretiert werden [130].

Unabhängig von der Einschätzung des Eintretens der alternativen Szenarien ist es aber letztlich entscheidend, ob der mit ihnen ermittelte Unsicherheitsbereich tatsächlich das Spektrum aller potentiellen zukünftigen Bevölkerungsentwicklungen abdeckt. In diesem Zusammenhang haben Analysen vergangener Bevölkerungsprognosen mit der Szenario-Technik oft einen zu engen Unsicherheitsbereich festgestellt [8], mit dem die Prognoseunsicherheit weder umfassend abgedeckt noch exakt quantifiziert werden kann.

Aus diesem Grund sind die so genannten probabilistischen Bevölkerungsprognosen entwickelt worden, die nun versuchen, das Spektrum aller potentiellen zukünftigen Bevölkerungsentwicklungen vollständig abzudecken, indem sie die Prognoseunsicherheit umfassend einfangen und exakt quantifizieren. Demnach ist das zunehmende Bewusstsein für die Prognoseunsicherheit ein Auslöser für die Weiterentwicklung der Methodik zur Bevölkerungsprognose.

Ein wesentlicher Ausgangspunkt für die Entwicklung probabilistischer Bevölkerungsprognosen besteht darin, Prognosefehler zu erkennen und Quellen ihrer Entstehung zu analysieren; so sind z.B. ex-post Fehleranalysen (vgl. Abschnitt 4.2.2 ab Seite 61) durchgeführt und Fehlerquellen beim Ablauf eines Prognoseprozesses (vgl. 4.1 ab Seite 52) zunehmend erkannt worden. Das Wissen um das Ausmaß und die Entstehung vergangener bzw. empirischer Prognosefehler kann dann mitunter zur Bestimmung der Prognosefehler neu zu erstellender Bevölkerungsprognosen eingesetzt werden.

Die Abbildung 3.3 stellt den generellen Ablauf einer probabilistischen Bevölkerungsprognose graphisch dar. Für jeden Modellparameter des Prognosemodells können verschiedene

Annahmenverläufe, die deren potentielle zukünftige Entwicklungen abbilden, generiert werden. Die Methode zur Generierung der Annahmen kann modell-basiert oder extern, frei wählbar sein. Für einen Prognosedurchlauf wird mittels einer Zufallsziehung ein Annahmenverlauf pro Modellparameter gezogen und in das Prognosemodell eingegeben. Die Wahl der Annahmenverläufe für die Modellparameter kann sowohl abhängig als auch unabhängig voneinander erfolgen. Der Simulator berechnet aus den zufällig gezogenen Annahmen einen Ergebnisverlauf pro Ergebnisgröße. Diese Prozedur wird  $n$  Male wiederholt, sodass nach  $n$  Prognosedurchläufen  $n$  Ergebnisverläufe pro Ergebnisgröße herauskommen. Diese  $n$  Ergebnisverläufe bilden eine Ergebnisverteilung. Die Häufigkeit, mit der ein Ergebnisverlauf eintritt, bestimmt seine Eintrittswahrscheinlichkeit.

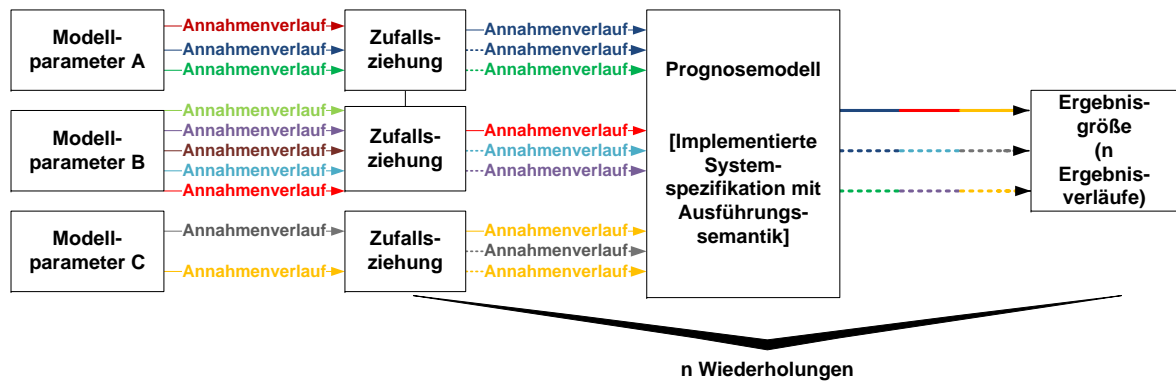


Abbildung 3.3.: Ablauf einer probabilistischen Bevölkerungsprognose. Pro Prognosedurchlauf wird ein Annahmenverlauf für jeden Modellparameter zufällig gezogen und in das Prognosemodell eingegeben, sodass daraus ein Ergebnisverlauf pro Ergebnisgröße berechnet wird. Beispielhaft wird im ersten Prognosedurchlauf der blaue Annahmenverlauf für Modellparameter A, der rote Annahmenverlauf für Modellparameter B und der orange Annahmenverlauf für Modellparameter C zufällig gezogen, sodass der Ergebnisverlauf blau-rot-orange, der sich aus den gezogenen Annahmenverläufen zusammensetzt, entsteht. Angedeutet sind zwei weitere Prognosedurchläufe mit entsprechenden gestrichelten, zufällig gezogenen Annahmen- und Ergebnisverläufen.

Der hier beschriebene Ablauf einer probabilistischen Bevölkerungsprognose richtet sich prinzipiell nach der simulativen Ausführung des PPPM (vgl. Teil II). Abweichungen zu anderen probabilistischen Bevölkerungsprognosemodellen können bei der Generierung der

Annahmen, der Wahl der Annahmenverläufe und bei der Bestimmung der Ergebnisverteilung (vgl. Abschnitt 4.2 ab Seite 57) auftreten.

Eine ausführliche Besprechung probabilistischer Ansätze zur Bevölkerungsprognose erfolgt in Kapitel 4.

### 3.2.3. Mikromodelle und Makromodelle zur Bevölkerungsprognose

Mikro- und Makromodelle zur Bevölkerungsprognose unterscheiden sich hinsichtlich ihrer Untersuchungseinheit. Mikromodelle betrachten einzelne Individuen, wohingegen Makromodelle den Zusammenschluss der Individuen, also die Bevölkerung im Ganzen, betrachten.

Seit der Vorlage des ersten Ansatzes zur Mikrosimulation von Orcutt [210] in der Mitte des 20. Jahrhunderts sind im Zuge der erhöhten und vermehrt verfügbaren Rechnerleistung zahlreiche neue Mikrosimulationsansätze entwickelt worden [293].

Gemeinhin wird bei einer Bevölkerungsprognose mit einem Mikromodell zunächst eine repräsentative Stichprobe aus der zu prognostizierenden Bevölkerung gezogen. Für jedes Individuum dieser Stichprobe wird in Abhängigkeit seiner zu betrachtenden Merkmale<sup>15</sup> ein Lebensverlauf, der durch den Verbleib und den Wechsel zwischen den zu betrachtenden Stati<sup>16</sup> gekennzeichnet ist, konstruiert. Ereignisse wie der Statuswechsel und der Verbleib in einem Status werden durch vorher festzulegende Eintrittswahrscheinlichkeiten in Abhängigkeit der zu betrachtenden Merkmale<sup>17</sup> bestimmt. Bei einer stochastischen Mikrosimulation wird z.B. mit der Monte-Carlo Simulation für jedes Individuum ein Zufallsexperiment für jedes Ereignis pro Projektionsintervall durchgeführt. Ist die generierte Zufallszahl, die generell zwischen null und eins liegen muss, größer als die Eintrittswahrscheinlichkeit des betroffenen Ereignisses, findet das Ereignis nicht statt. Das Ereignis findet hingegen statt, wenn die Zufallszahl kleiner als die Eintrittswahrscheinlichkeit des betroffenen Ereignisses ist. Über den gesamten Projektionshorizont werden mit den Zufallsexperimenten Lebensverläufe

---

<sup>15</sup>Zu betrachtende Merkmale können z. B. das Alter, das Geschlecht, die Kinderzahl, der Familienstand oder auch der Bildungsabschluss sein. Sie variieren mit der jeweiligen Fragestellung und den verfügbaren Daten.

<sup>16</sup>Zu betrachtende Stati können z. B. *kinderlos*, *1 Kind*, *2 Kinder*, *3 und mehr Kinder* und *tot* sein. Auch sie variieren wie die zu betrachtenden Merkmale nach der Fragestellung und den verfügbaren Daten.

<sup>17</sup>Je nach dem zu verwendenden Mikrosimulationsmodell können z. B. neben dem Alter und dem Geschlecht auch der aktuelle Status, die bisherige Statusabfolge und die Stati der anderen zu betrachtenden Individuen die Übergangsraten bestimmen.

einzelner Individuen simuliert. Die prognostizierte Bevölkerungszahl ergibt sich aus dem Aggregat der Individuen und deren anteiliger Hochrechnung auf die gesamte Bevölkerung in jedem Projektionsintervall. Neben der gesamten Bevölkerungszahl kann mit einem Mikromodell aber auch die Verteilung der Bevölkerung (nach verschiedenen Merkmalen) auf bestimmte Stati prognostiziert werden. Zudem können mit Mikromodellen Auswirkungen einzelner Veränderungen in unterschiedlichen gesellschaftlichen Bereichen<sup>18</sup> auf die Entwicklung individueller Lebensverläufe wie auch auf die Interaktion der Individuen untereinander dynamisch simuliert werden. Allerdings werden für derart detaillierte Prognosen auf der Mikroebene auch sehr viele, detaillierte Daten benötigt, die in der Form oftmals nicht verfügbar sind und deshalb häufig geschätzt werden müssen. Folglich entsteht durch den Versuch der detailgetreuen Abbildung der Realität auf der Individualebene mit einem Mikromodell vermehrt Unsicherheit [24, 293, 259, 260].

Bei einer Bevölkerungsprognose mit einem Makromodell wird hingegen die gesamte Bevölkerung auf der Basis eines demographischen Grundmodells wie der Kohorten-Komponenten Methode (vgl. Abschnitt 3.2.1 ab Seite 31) prognostiziert. Dabei wird die Bevölkerung durch demographische Ereignisse wie Geburten, Sterbefälle, Zuzüge und Fortzüge in Abhängigkeit von einigen Merkmalen wie z.B. dem Alter und dem Geschlecht fortgeschrieben.

Ein wesentlicher Unterschied zwischen Makro- und Mikromodellen besteht darin, dass bei einem Makromodell allen Personen einer Merkmalsgruppe, also z. B. allen Personen desselben Alters und desselben Geschlechts, dieselbe Wahrscheinlichkeit, ein demographisches Ereignis zu erfahren, unterstellt wird. Folglich wird die demographische Heterogenität [295] in Mikromodellen durch die Prognose demographischer Ereignisse auf der Individualebene<sup>19</sup> stärker berücksichtigt als in Makromodellen. Jedoch sind Mikromodelle durch den Einschluss mehrerer (individuell ausgeprägter) Erklärungsvariablen deutlich komplexer und es bleibt zu beweisen, ob sie tatsächlich genauere Prognoseergebnisse als Makromodelle erzielen können. Zumal durch die exakte Modellierung der Realität sehr detaillierte Daten benötigt werden und der Rechenaufwand bei der Ausführung eines Prognosemodells auf der Mikroebene enorm zunimmt [293].

---

<sup>18</sup>Eine Veränderung in einem gesellschaftlichen Bereich kann z. B. eine familienpolitische Maßnahme wie die Erhöhung des Kindergeldes sein.

<sup>19</sup>bzw. in Abhängigkeit zusätzlicher individueller Merkmale

Im Rahmen des internationalen Forschungsprojekts *MicMac — Bridging the micro-macro gap in population forecasting* werden erstmals ein Mikro- und ein Makromodell (Mic und Mac) zur Bevölkerungsprognose miteinander verbunden. Mic und Mac nutzen dieselben Übergangsraten oder *transition intensities* zur Prognose von Individual- bzw. Kohortenbiographien. Die Übergangsraten bestimmen in Abhängigkeit vom Alter, der Kalenderzeit, der Verweildauer im aktuellen Status und persönlichen Merkmalen, ob für eine Person ein Übergang von einem Status in einen anderen Status zu einem bestimmten Zeitpunkt stattfindet. Dadurch können z.B. die Annahmen zur Fertilität und Mortalität (sowohl in Mic als auch in Mac) nicht nur nach dem Alter und dem Geschlecht, sondern auch in Abhängigkeit weiterer persönlicher Merkmale variieren.<sup>20</sup> Außerdem soll das Verhalten der Personen (in Mic) nicht ausschließlich von ihren eigenen Merkmalen, sondern auch von ihrer Umwelt, also anderen Personen und deren Stati(abfolge) abhängen [91, 92, 308, 309, 310].

Generell sind Mikro- und Makromodelle in der Bevölkerungsprognose zwei alternative Ansätze, die letztlich beide die zukünftige Bevölkerungszahl und ihre Struktur nach bestimmten Merkmalen prognostizieren. Jedoch ist die Prognose der zukünftigen Bevölkerungsentwicklung bei Mikromodellen eher ein Nebenprodukt, wohingegen sich Makromodelle vornehmlich darauf spezialisiert haben.

#### 3.2.4. Strukturmodelle zur Bevölkerungsprognose

System- oder Strukturmodelle versuchen die Bevölkerung oder auch eine (oder mehrere) demographische Komponente(n) als abhängige Variable in einem komplexen Zusammenspiel diverser Einflussfaktoren zu prognostizieren. Dazu schließen sie neben demographischen auch nicht-demographische Erklärungsfaktoren wie beispielsweise aus der Ökonomie, der Umwelt, der Politik oder auch aus der Medizin zur Prognose der zukünftigen Bevölkerungsentwicklung ein. Oftmals sind System- oder Strukturmodelle, die eine demographische Komponente über verschiedene Erklärungsfaktoren prognostizieren, für eine Bevölkerungsprognose mit der Kohorten-Komponenten Methode (vgl. Abschnitt 3.2.1 ab Seite 31) verbunden [256].

System- oder Strukturmodelle versuchen zwar eine Bevölkerung umfassend zu prognostizieren, jedoch nimmt durch den Einschluss zahlreicher Erklärungsfaktoren die Komplexität

---

<sup>20</sup>In diesem Zusammenhang wird deshalb von *intra-cohort variation* oder auch von *cohort heterogeneity* gesprochen.

des zu verwendenden Prognosemodells zu. So muss für jeden Modellparameter, der einen Erklärungsfaktor für die zukünftige Bevölkerungsentwicklung repräsentiert, eine Annahme über dessen zukünftigen Verlauf getroffen werden. Allerdings wird auch die zukünftige Entwicklung eines jeden Modellparameters von diversen Faktoren beeinflusst. Je vielseitiger die zu betrachtenden Erklärungsfaktoren auf die zukünftige Bevölkerungsentwicklung demnach sind, desto höher ist die Anzahl der insgesamt bei einer Prognose direkt zu berücksichtigenden Faktoren. Folglich nimmt die Unsicherheit einer Prognose mit zunehmender Komplexität zu [63, 168]. Außerdem gibt es in der Demographie eine langanhaltende Diskussion über den Nutzen komplexer Bevölkerungsprognosemodelle hinsichtlich der Genauigkeit der Prognoseergebnisse (vgl. 8.2 ab Seite 161). In diesem Zusammenhang haben Vergleiche zu anderen Bevölkerungsprognosemodellen gezeigt, dass System- oder Strukturmodelle bisher keine genaueren Prognoseergebnisse hervorgebracht haben [30].

Nicht zu missachten ist jedoch deren zusätzlicher Erkenntnisgewinn. Anstatt ausschließlich die zukünftige Bevölkerungszahl und ihre Struktur z.B. nach Alter und Geschlecht zu prognostizieren, können mit System- oder Strukturmodellen auch Effekte aus verschiedenen Bereichen wie z.B. aus der Ökonomie oder der Umwelt auf die zukünftige Bevölkerungsentwicklung ermittelt werden. Insbesondere für diesen Zweck werden sie auch immer wieder für Bevölkerungsprognosen verwendet [168, 256]. Meadows et al. [187, 188, 189] haben z.B. ein populäres Strukturmodell vorgelegt, mit dem sie u.a. einen Bevölkerungsrückgang als Auswirkung global andauernder Wachstumstrends in der Industrialisierung, der Umweltverschmutzung, der Nahrungsmittelproduktion und im Ressourcenabbau aus der jeweiligen Gegenwart prognostiziert haben. Als Begründung führen Meadows et al. u.a. den verstärkten Einsatz von Maßnahmen zur Verbesserung der Stellung von Frauen auf, um ausgleichend in das Reproduktionsverhalten von Entwicklungsländern einzugreifen und damit (den Problemen) einer Überbevölkerung entgegenzuwirken. Allerdings stimmen die Bevölkerungsprognoseergebnisse nicht mit der tatsächlich eingetretenen Bevölkerungsentwicklung überein. Darüber hinaus sind der Versuch und die Wirkung der Einflussnahme auf das Reproduktionsverhalten anderer Gesellschaften umstritten [89].

Demnach eignen sich System -oder Strukturmodelle nicht vorrangig zur Bevölkerungsprognose an sich, sondern eher zur Ermittlung zukünftiger bevölkerungsdynamischer Effekte diverser Entwicklungen aus verschiedenen (z.B. ökonomischen oder politischen) Bereichen. Jedoch versuchen sie dabei die Realität detailliert durch eine Vielzahl vielfältiger Einflussfak-



toren abzubilden, sodass die Komplexität und demzufolge auch die Unsicherheit der Prognose (unverhältnismäßig) steigen.

### 3.2.5. *Projections* und *Forecasts*

Bevölkerungsprognosen können nicht nur im Hinblick auf die Eintrittswahrscheinlichkeit ihrer Ergebnisse oder im Hinblick auf ihre Untersuchungseinheit oder die Vielfalt der einzuschließenden Erklärungsfaktoren, sondern eben auch hinsichtlich ihrer eigentlichen Zielstellung unterschieden werden. So werden in der englischen Fachliteratur *projections* von *forecasts* unterschieden [137, 256]. Dabei sind unter *projections* Modellrechnungen und unter *forecasts* echte Vorausschätzungen zu verstehen.

Projektionen zeichnen sich dadurch aus, dass sie nicht die Prognose des wahrscheinlichsten Verlaufs der zukünftigen Bevölkerungsentwicklung zum Ziel haben, sondern vielmehr bevölkerungsdynamische Effekte bestimmter Entwicklungen einer (oder mehrerer) demographischer Komponente(n) aufzeigen sollen. Dabei können generell Zielprojektionen von Wirkungsprojektionen unterschieden werden. Zielprojektionen ermitteln den potentiellen Verlauf einzelner Komponenten, um ein vorgegebenes Ergebnis wie z.B. eine bestimmte Bevölkerungszahl in einem vorher festgelegten Zeitraum zu erzielen. Hingegen werden bei Wirkungsprojektionen die Effekte bestimmter Entwicklungen für die demographischen Komponenten auf die zukünftige Bevölkerungsentwicklung analysiert. Aufgrund dieser Vorgehensweise werden bei Zielprojektionen Handlungsbedarf und bei Wirkungsprojektionen Handlungskonsequenzen aufgedeckt [198].

Bevölkerungsprognosen im Sinne von Vorausschätzungen versuchen den wahrscheinlichsten Verlauf der zukünftigen Bevölkerungsentwicklung zu ermitteln. Dafür treffen sie für die einzelnen demographischen Komponenten realistische Annahmen, die, zumindest zum Zeitpunkt der Prognoseerstellung, die höchste Wahrscheinlichkeit haben, tatsächlich einzutreffen.

Projektionen können ausschließlich durch Berechnungsfehler ungenaue Ergebnisse generieren, wohingegen Bevölkerungsprognosen im Sinne echter Vorausschätzungen zudem als fehlerhaft bewertet werden können, wenn die prognostizierten mit den später tatsächlich eingetretenen Werten nicht übereinstimmen. Um das Ausmaß der fehlenden Übereinstimmung zwischen prognostizierten und beobachteten Werten zu quantifizieren, können diverse Fehlermaße (vgl. Abschnitt 8.1 ab Seite 147) berechnet werden.

Die hier beschriebene Unterscheidung der verschiedenen Prognosearten ist nicht universal gültig. Neben *projections* und *forecasts* können z. B. auch noch *predictions* und *estimations* im Bereich der Prognosen unterschieden werden. In dieser Arbeit sollen jedoch all diese verschiedenen Begriffe synonym als Prognose oder auch als Projektion verwendet werden.

### 3.2.6. Zusammenfassung

Derzeit gibt es eine Vielfalt bestehender Ansätze zur Bevölkerungsprognose, die sich u. a. in der Berücksichtigung der Prognoseunsicherheit, der Untersuchungseinheit, der einzuschließenden Erklärungsfaktoren und der eigentlichen Zielstellung unterscheiden.

Nichtsdestotrotz basieren diese modernen Ansätze zur Bevölkerungsprognose meist auf einer Variante der Kohorten-Komponenten Methode, die eine Bevölkerung nach Alter (und möglicherweise weiteren Merkmalen) mit Annahmen zu demographischen Ereignissen aus den Bereichen der Fertilität, Mortalität und Migration prognostiziert.

Da der Fokus dieser Arbeit auf die adäquate Berücksichtigung von Prognoseunsicherheit ausgerichtet ist, sollen nachfolgend in Kapitel 4 Unsicherheitsquellen wie auch bestehende probabilistische Makromodelle zur Bevölkerungsprognose näher beleuchtet werden.

## 4. Überblick über probabilistische Bevölkerungsprognosemodelle

Eine probabilistische Bevölkerungsprognose versucht wahrscheinliche Zukunftsaussagen über das komplexe System einer Bevölkerung zu treffen. Dabei bestimmt das Zusammenwirken von Fertilität, Mortalität und Migration die zukünftige Bevölkerungszahl und -struktur. Da jedoch die Bevölkerung aus einer Vielzahl komplexer Individuen besteht, deren Verhalten — gerade auch in Bezug auf das Eintreten demographischer Ereignisse — auf individuellen Entscheidungsprozessen basiert, können diese aufgrund ihrer Privatheit und Instabilität oft nur unvollkommen antizipiert werden. Aus diesem Grund kann die Unsicherheit von Prognosen auch als wesentliche Eigenschaft dieser verstanden werden, sodass bei Bevölkerungsprognosen auch von einer *prognoseimmanenten Unsicherheit* gesprochen werden kann.<sup>1</sup>

Neben der prognoseimmanenten Unsicherheit gibt es in jedem Arbeitsschritt eines Prognoseprozesses (vgl. Abschnitt 2.2 ab Seite 18) vielfältige Fehler- und Unsicherheitsquellen, die die Prognoseergebnisse verzerren können. Bei der Entwicklung eines Bevölkerungsprognosemodells ist vornehmlich die Auswahl geeigneter Modellparameter wie auch die Generierung ihrer Annahmenwerte mit Unsicherheit behaftet. Darüber hinaus können besonders bei der Sammlung relevanter Daten oder auch bei der Implementierung und Ausführung eines Prognosemodells Fehler auftreten. Derlei Fehler- und Unsicherheitsquellen werden ausführlich im Abschnitt 4.1 ab Seite 52 erläutert.

Probabilistische Bevölkerungsprognosen versuchen, die bekannten Fehler in jedem Arbeitsschritt zu vermeiden und die verbleibende Unsicherheit mit geeigneten Methoden einzufangen und zu quantifizieren. Dazu ermitteln sie anstatt nur *eines* Ergebnisverlaufs eine Ergebnisverteilung aus *mehreren* Ergebnisverläufen pro Ergebnisgröße. Mit dieser Ergebnisverteilung kann dann die Prognoseunsicherheit quantifiziert werden, indem jedem Prognoseergebnis ei-

---

<sup>1</sup>So schätzt z. B. Stoto [272] Prognosen als *inherently inaccurate* ein und Lee [152] bringt Prognosen mit einer *intrinsic uncertainty* in Verbindung.

ne Eintrittswahrscheinlichkeit zugeordnet wird (vgl. Abschnitt 3.2.2 auf Seite 41). Generell sollen mit probabilistischen Bevölkerungsprognosen nicht nur sichere, sondern auch genaue Prognoseergebnisse generiert werden. Allerdings bewirken diese beiden Ziele gegensätzliche Effekte: Während eine sichere probabilistische Prognose eher eine breitere Ergebnisverteilung erzeugt, erzeugt eine genauere probabilistische Prognose eher eine schmalere Ergebnisverteilung [130]. Ein Überblick über die verschiedenen bestehenden Ansätze zur probabilistischen Bevölkerungsprognose wird im Abschnitt 4.2 ab Seite 57 gegeben. Darüber hinaus werden exemplarisch einige Vertreter für die unterschiedlichen Kategorien probabilistischer Bevölkerungsprognoseansätze vorgestellt.

### 4.1. Fehler- und Unsicherheitsquellen beim Bevölkerungsprognoseprozess

Jeder Arbeitsschritt eines Bevölkerungsprognoseprozesses (vgl. Abschnitt 2.2 auf Seite 18) enthält diverse Fehler- und Unsicherheitsquellen. Die Kenntnis dieser Fehler- und Unsicherheitsquellen kann dazu führen, Fehler zu vermeiden und die verbleibende Prognoseunsicherheit angemessen zu berücksichtigen, indem sie noch besser eingefangen und z. T. quantifiziert wird.<sup>2</sup>

Bei der Formulierung der Frage- bzw. Zielstellung einer Bevölkerungsprognose können unaufgedeckte, unterschiedlich verstandene Fachtermini zu einem Missverständnis zwischen Auftraggebern, Erstellern und Nutzern einer Bevölkerungsprognose führen.

Ebenso enthält die Entwicklung eines Prognosemodells bei der Wahl der relevanten Modellparameter und einer für sie geeigneten Abstraktionsebene Unsicherheit. Ein Prognosemodell versucht grundsätzlich die in der Realität auftretende Bevölkerungsbewegung möglichst wirklichkeitsgetreu abzubilden. Da jedoch nicht alle Einzelheiten bis ins letzte Detail modelliert werden können, müssen vereinfachend wesentliche Einflussfaktoren erkannt und als Modellparameter festgelegt werden. Allerdings können bei dieser Auswahl nicht nur unnötige, sondern auch notwendige Einflussfaktoren und Ergebnisgrößen ausgeschlossen werden. Neben der Wahl der relevanten Modellparameter ist auch der zu wählende Grad der Disaggregation bei jedem Modellparameter entscheidend. So kann ein zu geringer Dis-

---

<sup>2</sup>Eine Beschreibung verschiedener Fehlerquellen einer Bevölkerungsprognose ist u. a. bei Lee [156] und Alho und Spencer [10] zu finden.

aggregationsgrad einiger Modellparameter dazu führen, dass die zentrale Fragestellung nur ungenügend beantwortet werden kann. Beispiele für das Entstehen zusätzlicher Unsicherheit durch die Wahl eines zu geringen Disaggregationsgrades einiger Modellparameter sind der Verzicht auf eine Einzelalters- und Geschlechtsstruktur oder auch die Betrachtung von Netto- statt Bruttomigration. So entstehen erst durch die Unterscheidung von Einzelaltern und Geschlechtern relativ homogene Subpopulationen, die sich aus Personen zusammensetzen, deren demographisches Verhalten bestimmte Regelmäßigkeiten bzw. Ähnlichkeiten aufweist, sodass die Treffsicherheit von Annahmen gesteigert werden kann. Zudem werden bei der alleinigen Betrachtung von Nettomigration bevölkerungsdynamische Effekte, die aufgrund der unterschiedlichen Alters- und Geschlechtsstruktur der Zu- und Fortzüge entstehen, nicht berücksichtigt, obwohl sie einen entscheidenden Einfluss auf die Entwicklung der Gesamtbevölkerung haben können (vgl. Abschnitt 3.2.1 ab Seite 31). Folglich führen der Ausschluss notwendiger und der Einschluss unnötiger Modellparameter wie auch die Wahl eines ungeeigneten Abstraktionsniveaus für die Modellparameter zu einer erhöhten Prognoseunsicherheit [169, 307].

Darüber hinaus kann Unsicherheit durch verschiedene Arten der Modellmisspezifikation entstehen. Einerseits können die festzulegenden definitorischen Zusammenhänge der einzubauenden Modellparameter Fehler aufweisen. Andererseits kann ein Prognosemodell generell für die Beantwortung der eigentlichen Fragestellung ungeeignet sein. Dieser Fall trifft eher bei der Verwendung eines bereits bestehenden Prognosemodells als bei der Entwicklung eines neuen Prognosemodells ein. So ist beispielsweise ein Makromodell ungeeignet, Individualbiographien zu prognostizieren (vgl. Abschnitt 3.2.3 ab Seite 45).

Bei der Implementierung eines Simulators zur computergestützten Ausführung eines Prognosemodells stellt ein Programmierfehler in der Software eine Fehlerquelle dar, die aufgrund einer fehlerhaften Ausführung zur Berechnung verzerrter Ergebnisse führen kann. Um diese Fehlerquelle auszuschalten, kann der gleiche Simulator für dasselbe formale Prognosemodell von verschiedenen Personen in derselben Computersprache oder aber in verschiedenen Computersprachen geschrieben werden. Die Replikation des Simulators ermöglicht den Vergleich der jeweils mit demselben Prognosemodell berechneten Prognoseergebnisse, sodass Programmierfehler aufgedeckt und eliminiert werden können. In diesem Zusammenhang ist auch die Verifikation eines implementierten Simulators zur Ausführung eines Prognosemodells zu nennen, bei der überprüft wird, ob der Simulator exakt die formale

Spezifikation des entsprechenden Prognosemodells enthält. Andernfalls können fehlerhafte Prognoseergebnisse generiert werden.

Fehler in der zu verwendenden Hardware können ebenfalls verzerrte Prognoseergebnisse erzeugen. Sie können — wenn überhaupt — durch die Nutzung unterschiedlicher Computer umgangen werden. Neben der Hardware können aber auch einzelne Bestandteile der Software unvollkommen und damit fehlerbehaftet sein. Ein Beispiel hierfür stellt die Verwendung eines Zufallszahlengenerators dar. So ist es technisch und methodisch sehr kompliziert, echte Zufallszahlen zu generieren (vgl. Abschnitt 10.4.1 ab Seite 216).

Daneben kann Unsicherheit bei der Sammlung relevanter Daten für eine Prognose entstehen, wenn entscheidende Daten nicht verfügbar oder fehlerhaft sind. So können Daten für bestimmte Modellparameter einer demographischen Komponente überhaupt nicht erhältlich sein, Daten ungenügend disaggregiert vorliegen oder nur in unzureichender Menge für eine bestimmte (vornehmlich statistische) Methode (z. B. als zu kurze historische Zeitreihe) vorhanden sein. Außerdem können relevante, verfügbare Daten Fehler enthalten, die u. a. durch eine fehlerhafte Erhebung oder Dokumentation entstehen können.

Relevante, jedoch nicht verfügbare Daten können dazu führen, dass ein komplexes Prognosemodell, das diese Daten benötigt, verworfen werden muss. Anstattdessen wird dann ein einfacheres Prognosemodell herangezogen, das (im Vergleich zu einem komplexeren Prognosemodell) Modellparameter mit einem niedrigeren Disaggregationslevel wie auch einfachere Methoden zur Generierung der Annahmen (vgl. Kapitel 5, 6 und 7) nutzt. Insbesondere unter instabilen Bedingungen wirkt sich dieses Vorgehen nachteilig aus, da komplexe Prognosemodelle unerwartete Veränderungen besser einfangen können als einfache Prognosemodelle. Deshalb nimmt mit der Verwendung einfacherer Prognosemodelle die Unsicherheit gerade bei Langzeitprognosen unter instabilen Bedingungen zu [307] (vgl. Abschnitt 8.2 ab Seite 161).

Im Vergleich dazu können fehlerhafte (Ausgangs-)Daten die Prognoseergebnisse verzerren, da sie die Gefahr erhöhen, fehlerhafte Annahmen für die einzelnen Modellparameter zu generieren. Fehler in den Originaldaten sind sehr tückisch, da sie schwer festzustellen und kaum zu korrigieren sind (mit Ausnahme einer erneuten Erhebung).

Darüber hinaus ist der Arbeitsschritt der Annahmengenerierung insbesondere durch diskontinuierliche Entwicklungen einzelner Modellparameter mit Prognoseunsicherheit behaftet. Unerwartete Ereignisse<sup>3</sup> oder auch die erschwerte Prognostizierbarkeit von demographischem Verhalten, das auf individuellen Entscheidungsprozessen basiert, fördern die Prognoseunsicherheit nahezu unkontrollierbar.

Außerdem geben Prognosemodelle häufig eine (modell-basierte) Methode zur Generierung der Annahmen für die Modellparameter vor. Dadurch können die Annahmenwerte der Modellparameter in ihrem Niveau wie auch in ihrem Verlauf übermäßig stark eingeschränkt werden, sodass sie nicht alle potentiellen zukünftigen Entwicklungen adäquat berücksichtigen können. So bestimmt z. B. die Art der vorzugebenden modell-basierten Methode(n), welche Einflussfaktoren bzw. Informationsquellen bei der Annahmengenerierung berücksichtigt werden. Während einige Methoden ausschließlich in der Vergangenheit beobachtete Werte einbeziehen, berücksichtigen andere Methoden zusätzlich Theorien zur zukünftigen Bevölkerungsentwicklung aus der Demographie und benachbarten Wissenschaftsdisziplinen wie auch Ergebnisse aus themenrelevanten Erhebungen via Expertenmeinung. Die Nutzung möglichst aller relevanten Quellen zur Generierung der Annahmen für die Modellparameter eines Prognosemodells mit möglichst vielfältigen Methoden könnte hingegen die Prognoseunsicherheit angemessen einfangen (vgl. Abschnitt 10.2.1 ab Seite 200).

Zudem sollten die mit verschiedenen Methoden zu generierenden Annahmen für die Modellparameter demographisch erklärbar sein. Denn nur, wenn verschieden modellierte Annahmenverläufe demographisch interpretierbar sind, kann auch das Ausmaß der Abdeckung der potentiellen, realistischen zukünftigen Entwicklungen eingeschätzt werden. Demographisch nicht erklärbare Annahmenverläufe können beispielsweise mit statistischen Methoden generiert werden, deren Parameter selbst nicht demographisch interpretierbar sind. Ein Beispiel ist in diesem Zusammenhang das komplexe Mortalitätsgesetz von Rogers und Little [239] (vgl. 5.2.1 ab Seite 90).

Eine weitere Unsicherheitsquelle betrifft die Entscheidung, ob die Entwicklung von Modellparametern unterschiedlicher demographischer Komponenten voneinander (un)abhängig ist.

---

<sup>3</sup>Unerwartete Ereignisse können sich beispielsweise in Bereichen wie der Politik, der Medizin, der Wirtschaft oder auch in der Umwelt ereignen. Sie beeinflussen die zukünftige Bevölkerungsentwicklung maßgeblich, können von einem Prognostiker aber nur schwer antizipiert werden.

Wenn sie abhängig voneinander modelliert werden, ist es zudem fraglich, ob ein Ansteigen der Annahmenwerte der einen Komponente auch ein Ansteigen der Annahmenwerte der anderen Komponente(n) nach sich zieht und umgekehrt (vgl. Abschnitt 10.2.2 ab Seite 203). In diesem Zusammenhang gehen u. a. die Theorie des *Parallelismus* von Bertillon oder auch der *Demographische Übergang* von Thompson [280] und später auch von Notestein [200] von einer (ursächlich) abhängigen, (teils) parallelen Entwicklung der Geburten- und Sterberate über die Zeit aus. Jedoch haben diese Behauptungen allenfalls Modellcharakter, da bereits in der Vergangenheit gegenläufige empirische Entwicklungen der Fertilität und Mortalität nachgewiesen werden konnten [71]. Aufgrund dieser Ungewissheit über den abhängigen oder unabhängigen Verlauf der demographischen Komponenten gibt es sowohl Unsicherheit bei der einen als auch bei der anderen Modellierungsvariante.

Weitere Fehlerquellen stellen bei der Ausführung einer Prognose Hardware-, Software- und Rechenfehler dar.

Bei der Überprüfung der berechneten Prognoseergebnisse ob ihrer Eignung zur Beantwortung der Fragestellung kann Unsicherheit entstehen, wenn diese aufgrund erneuter Missverständnisse verworfen werden.

Da bei der Bewertung von Prognoseergebnissen Fehlermaße (vgl. Abschnitt 8.1 ab Seite 147) zur Quantifizierung der Abweichungen zwischen den prognostizierten und den tatsächlich eingetretenen Werten eingesetzt werden, können hierbei Ungenauigkeiten durch die Verwendung eines ungeeigneten Fehlermaßes oder auch durch dessen fehlerhafte Berechnung auftreten.

Abschließend birgt die Verwendung der Prognoseergebnisse durch den Endnutzer die Gefahr der Fehlinterpretation.

Die adäquate Berücksichtigung der hier aufgeführten, vielfältigen Unsicherheits- und Fehlerquellen kann die Vorhersagekraft einer Bevölkerungsprognose verbessern. Inwiefern bestehende probabilistische Ansätze zur Bevölkerungsprognose diese Unsicherheitsquellen berücksichtigen, wird u. a. im folgenden Abschnitt 4.2 erläutert.



## 4.2. Systematik probabilistischer Ansätze zur Bevölkerungsprognose

Es gibt in der wissenschaftlichen Literatur diverse Vorschläge zur Einteilung probabilistischer Bevölkerungsprognoseansätze, die sich u. a. in der Terminologie und der Anzahl der zu differenzierenden Methoden unterscheiden. Die Systematik, die in dieser Arbeit verwendet werden soll, lehnt sich u. a. an die Arbeiten von Pflaumer [215], Smith et al. [256], Lee [156], Tuljapurkar et al. [282], Long [169, 168] und Keilman und Pham [130, 132] an.<sup>4</sup>

Wie die Abbildung 4.1 zeigt, können probabilistische Bevölkerungsprognoseansätze zunächst anhand des Aggregationsniveaus des Prognosegegenstands und der zu berücksichtigenden Modellparameter unterschieden werden. So kann z. B. die Bevölkerung entweder *aggregiert* oder *disaggregiert* nach verschiedenen Merkmalen wie dem Alter und dem Geschlecht prognostiziert werden. Daneben kann der Prognosegegenstand entweder ausschließlich über sich selbst oder aber unter Einschluss diverser Modellparameter prognostiziert werden. Mehrere Modellparameter werden beispielsweise bei der Kohorten-Komponenten Methode (vgl. Abschnitt 3.2.1 ab Seite 31) berücksichtigt. Weiterhin können probabilistische Bevölkerungsprognoseansätze dahingehend unterschieden werden, ob sie keine, eine oder mehrere modell-basierte Methoden zur Annahmengenerierung für die einzelnen Modellparameter vorgeben. Vorzugebende Methoden zur Annahmengenerierung können z. B. der Expertenmeinung, der Extrapolation, der Fehlerfortschreibung oder auch einem Methodenmix entsprechen. Bei einer Extrapolation wird ausschließlich die vergangene empirische Entwicklung des Prognosegegenstands oder auch einzelner Modellparameter mit einem bestimmten statistischen Verfahren in die Zukunft fortgeschrieben, wohingegen bei der Expertenmeinung zusätzliche Informationen z. B. inform von Theorien und Erkenntnissen aus Befragungen bei der Generierung der Annahmen berücksichtigt werden können. Im Vergleich dazu werden bei der Fehlerfortschreibung die Fehler vergangener Bevölkerungsprognosen als Unsicherheitsbereich auf die Bevölkerungsprognosen der Gegenwart übertragen. Neben den bereits genannten Merkmalen können probabilistische Bevölkerungsprognoseansätze außerdem dahingehend unterschieden werden, ob sie auf der Makro- oder auf der Mikroebene (vgl. Abschnitt 3.2.3 ab Seite 45) prognostizieren, ob die Projektionsintervalle diskret oder kontinuierlich sind und ob die Prognoserechnung analytisch oder simulativ ausgeführt wird.<sup>5</sup>

---

<sup>4</sup>Darüber hinaus haben beispielsweise Alho und Spencer [9], Ahlburg [2], Smith [254], Rogers [238], Smith und Sincich [255] einfachere Einteilungen für verschiedene Bevölkerungsprognoseansätze erarbeitet.

<sup>5</sup>Die Liste der hier aufgeführten wesentlichen Merkmale zur Unterscheidung probabilistischer Bevölkerungs-

#### 4. Überblick über probabilistische Bevölkerungsprognosemodelle

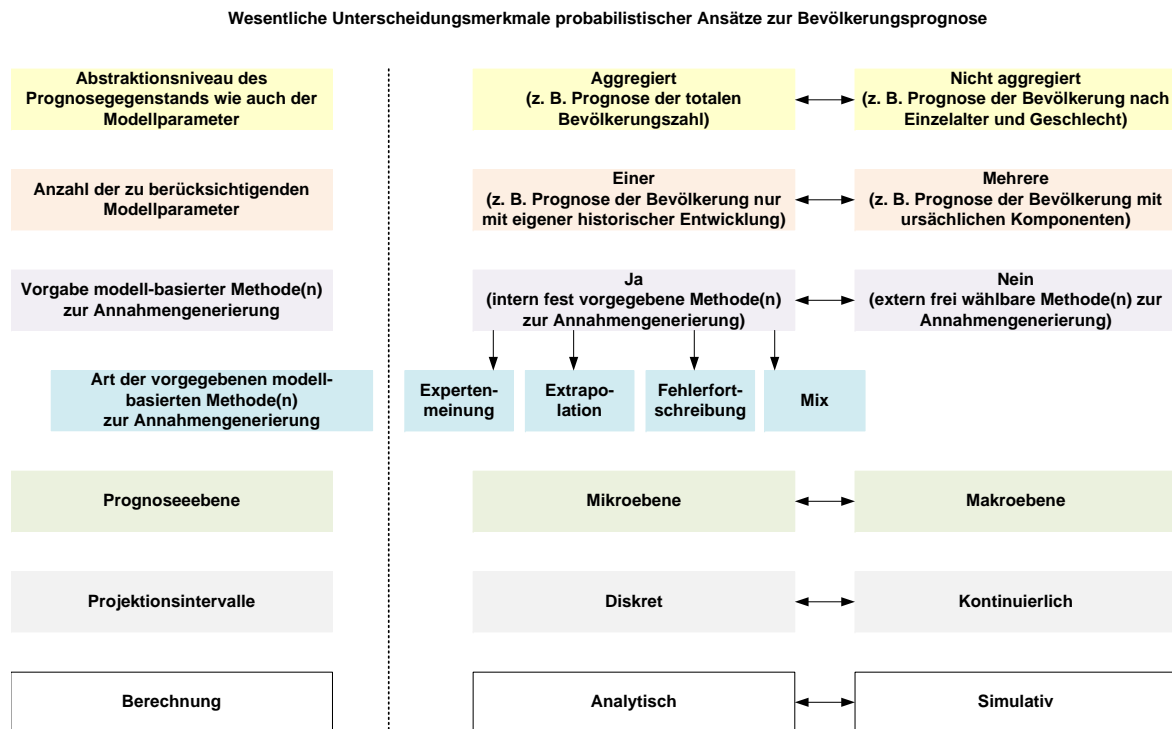


Abbildung 4.1.: Wesentliche Unterscheidungsmerkmale probabilistischer Ansätze zur Bevölkerungsprognose; den verschiedenen Merkmalen (links) werden ihre entsprechenden Ausprägungen (rechts) zugeordnet.

Die hier aufgeführten Unterscheidungsmerkmale errichten quasi einen Dimensionsraum, in dem die bestehenden Ansätze zur probabilistischen Bevölkerungsprognose hinsichtlich häufig auftretender Merkmalskombinationen systematisch angeordnet werden können.

Nachfolgend werden exemplarisch Ansätze zur probabilistischen Bevölkerungsprognose auf der Makroebene vorgestellt, die die totale Bevölkerungszahl extrapolieren (vgl. Abschnitt 4.2.1) und die Fehler vergangener auf aktuelle Prognosen übertragen (vgl. Abschnitt 4.2.2); außerdem werden in den Abschnitten 4.2.3 und 4.2.4 fortgeschrittenere Ansätze präsentiert, die eine Bevölkerung disaggregiert (nach Alter und Geschlecht) unter Berücksichtigung de-

---

prognoseansätze ist nicht erschöpfend. Beispielsweise können probabilistische Bevölkerungsprognoseansätze auch hinsichtlich ihrer Korrelationsstruktur zwischen verschiedenen Modellparametern unterschieden werden. Allerdings reichen diese wesentlichen Merkmale aus, um die bestehenden probabilistischen Bevölkerungsprognoseansätze nach entscheidenden Charakteristika zu unterscheiden und zu systematisieren.

mographischer Komponenten extrapolieren bzw. unter zusätzlichem Einsatz von Expertenmeinung simulativ berechnen; abschließend wird im Abschnitt 4.2.5 noch ein Ansatz näher erläutert, der verschiedene Methoden zur Generierung der Annahmen verbindet, um eine Bevölkerung umfassend prognostizieren zu können.

##### 4.2.1. Extrapolation der totalen Bevölkerungszahl

Relativ simple Bevölkerungsprognosen, die lediglich die historische Bevölkerungsentwicklung mit einem komplexen Zeitreihenmodell<sup>6</sup> extrapolieren, sind u. a. von Saboia [242] und Pflaumer [217] genutzt worden.

###### Der Ansatz von Saboia

Saboia [242] ist 1974 einer der ersten gewesen, der ein  $\text{ARIMA}(p,d,q)$ -Zeitreihenmodell aus der historischen Bevölkerungsentwicklung schätzt und zur Bevölkerungsprognose nutzt. Zur Prognose der schwedischen Bevölkerung bis zum Zieljahr 1980 schätzt Saboia mehrere  $\text{ARIMA}(p,d,q)$ -Zeitreihenmodelle auf der Basis der vergangenen schwedischen Bevölkerungsentwicklung von 1780 bis 1970 mit Projektionsintervallen der Länge von 5 Jahren. Um aus den Originaldaten der Basisperiode eine stationäre Zeitreihe zu erhalten, berechnet Saboia sowohl die Differenzen ersten als auch die Differenzen zweiten Grades. Die anschließende Analyse der Autokorrelationsfunktion und der partiellen Autokorrelationsfunktion lassen Saboia ein  $\text{ARIMA}(0,2,1)$ -Zeitreihenmodell favorisieren, mit dem er die schwedische Bevölkerung auf 8,586 Millionen Personen im Zieljahr 1980 prognostiziert.

###### Der Ansatz von Pflaumer

Pflaumer [217] prognostiziert 1992 die totale US-amerikanische Bevölkerung mit einem  $\text{ARIMA}(p,d,q)$ -Zeitreihenmodell bis zum Zieljahr 2088. Zur Schätzung eines geeigneten  $\text{ARIMA}(p,d,q)$ -Zeitreihenmodells nutzt er die historische Bevölkerungsentwicklung von 1900 bis 1988 als Basisdaten. Um die Originalwerte der Basisperiode in eine stationäre Zeitreihe umzuwandeln, bildet Pflaumer die Differenzen zweiten Grades. Anschließend kann er die Prozesse der Autoregression und des gleitenden Durchschnitts anhand der Autokorrelationsfunktion und der partiellen Autokorrelationsfunktion ermitteln. Da die Autokorrelationsfunktion

---

<sup>6</sup>Komplexe Zeitreihenmodelle sind von Box und Jenkins [36] in der zweiten Hälfte des 20. Jahrhunderts entwickelt worden. Sie können zur Prognose aller demographischen Prozesse, also zur Prognose der Mortalität, der Fertilität, der Migration wie auch zur Prognose der Bevölkerung verwendet werden.

betragsmäßig exponentiell fällt und die partielle Autokorrelationsfunktion zwei Spitzen aufweist, schließt Pflaumer auf ein ARIMA(2,2,0)-Modell. Pflaumer prognostiziert mit diesem ARIMA(2,2,0)-Modell, dass die US-amerikanische Bevölkerung im Zieljahr 2088 mit einer Wahrscheinlichkeit von 95 Prozent zwischen 265 Millionen und 712,6 Millionen Personen liegt und um den Mittelwert von 488,8 Millionen Personen schwankt. Den Vorteil einfacher Bevölkerungsprognosemodelle sieht Pflaumer zwar darin, dass sie oftmals ähnlich genaue Prognoseergebnisse erzielen wie komplexe Bevölkerungsprognosemodelle (vgl. Abschnitt 8.2 ab Seite 161). Allerdings befürwortet er auch die Weiterentwicklung komplexer Bevölkerungsprognosemodelle, indem demographisches Wissen zunehmend berücksichtigt und dadurch letztendlich die Prognosekraft erhöht wird.

### **Zusammenfassung**

Probabilistische Ansätze, die zur Prognose der totalen Bevölkerung lediglich die vergangene Bevölkerungsentwicklung extrapolieren, können zwar bei einer stabilen Entwicklung zufällig genaue Prognoseergebnisse erzeugen. Allerdings können sie eine instabile Bevölkerungsentwicklung nur ungenügend einfangen, da sie nicht alle verfügbaren (demographischen) Informationen nutzen, pauschal den in der Basisperiode enthaltenen Entwicklungstrend fortschreiben und die Prognoseunsicherheit durch die Konstruktion ihrer Konfidenzintervalle unterschätzen.

Um eine instabile Bevölkerungsentwicklung angemessen einfangen zu können, kann nicht einfach die historische Bevölkerung mit dem in der Basisperiode beobachteten Trend pauschal fortgeschrieben werden, da dadurch etwaige zukünftige Trendumkehrungen weder erkannt noch modelliert werden können. Anstattdessen sollte die Bevölkerungsentwicklung in Abhängigkeit ihrer demographischen Komponenten (also der Fertilität, Mortalität und Migration) jeweils unter Berücksichtigung spezieller Einflussfaktoren (ursächlich) prognostiziert werden, sodass die Auslöser instabiler Entwicklungen erkannt und (mit fortgeschritteneren Methoden) eingefangen werden können. Darüber hinaus bestimmt die Vorgabe eines komplexen Zeitreihenmodells bei den einfachen Bevölkerungsprognosemodellen von Saboia [242] und Pflaumer [217] nicht nur die Verlaufsstruktur der zukünftigen Bevölkerungsentwicklung, sondern auch die einzuschließenden Unsicherheitsquellen bei der Berechnung der Konfidenzintervalle. Da die Konfidenzintervalle über die Punktprognose und die Abweichung der mit dem Zeitreihenmodell geschätzten von den tatsächlich eingetroffenen Werten bestimmt werden, bleiben weitere Fehler- und Unsicherheitsquellen (vgl. Abschnitt

4.1 ab Seite 52) unberücksichtigt, sodass die Prognoseunsicherheit unterschätzt werden kann.

Folglich eignen sich einfache probabilistische Bevölkerungsprognoseansätze, die die totale Bevölkerung extrapolieren, lediglich für kurze Projektionshorizonte, in denen die Annahme stabiler Bedingungen noch einigermaßen plausibel ist. Außerdem können komplexe Zeitreihenmodelle nur zur Schätzung und Prognose einer Bevölkerung verwendet werden, für die eine genügend lange, konsistente Zeitreihe mit fehlerfreien Daten vorliegt.

#### 4.2.2. Fehlerfortschreibung

Eine ganz andere Vorgehensweise zeichnet die Ansätze mit Fehlerfortschreibung aus, die quasi deterministischen Bevölkerungsprognosen Unsicherheitsbereiche zuordnen, indem sie empirische Fehler vergangener auf aktuelle Bevölkerungsprognosen übertragen.

#### Der Ansatz von Keyfitz

Keyfitz [136] hat 1981 einen Ansatz zur Bestimmung eines Unsicherheitsbereichs für deterministische Bevölkerungsprognosen vorgelegt. Die Grundidee besteht darin, die Abweichungen bzw. Fehler<sup>7</sup> vergangener Bevölkerungsprognosen auf in der Gegenwart zu erstellende deterministische Bevölkerungsprognosen zu übertragen, sodass für sie daraus ein Unsicherheitsbereich wie bei einer probabilistischen Bevölkerungsprognose bestimmt werden kann.

Die genaue Auswertung von 810 Bevölkerungsprognosen von der UN ab dem Ende der 1950er zeigt, dass die prognostizierte jährliche Wachstumsrate einer Bevölkerung mit einer Wahrscheinlichkeit von 67 Prozent im Durchschnitt<sup>8</sup> zwischen  $+/- 0,4$  Prozentpunkten schwankt. Zur Bestimmung dieses Wertes gewichtet Keyfitz die Fehler weniger lang zurückliegender Bevölkerungsprognosen stärker, um die festgestellte Verringerung der Prognosefehler über die Zeit entsprechend zu berücksichtigen. Darüber hinaus ist die Höhe des durchschnittlichen Fehlers abhängig von der Höhe der prognostizierten jährlichen

---

<sup>7</sup>Ein Fehler entspricht dabei der Abweichung der prognostizierten von den tatsächlich eingetroffenen Bevölkerungswerten (vgl. Abschnitt 8.1 ab Seite 147).

<sup>8</sup>Keyfitz hat zur Bestimmung des durchschnittlichen Fehlers den *Root Mean Square Error* (vgl. Formel 8.5 auf Seite 151) verwendet, da er u. a. durch das Quadrieren der Abweichung ein größeres Gewicht auf größere Fehler legt.

Wachstumsrate einer Bevölkerung: Je höher das Bevölkerungswachstum ist, desto höher ist auch der Prognosefehler.<sup>9</sup> Daneben weist Keyfitz darauf hin, dass die Annahme des konstanten Fehlers von  $\pm 0,4$  Prozentpunkten nur für kurze Projektionshorizonte der Länge von 5 bis 20 Jahren valide ist.

Keyfitz ermöglicht mit diesem Ansatz die Berechnung eines 67 Prozent Konfidenzintervalls für deterministische Bevölkerungsprognosen auf der Basis vergangener Prognosefehler. Allerdings ist die Verwendung vergangener Prognosefehler nur gültig, wenn sich weder die Prognosemethoden verbessert, noch die Bevölkerungsbewegung grundlegend verändert haben. Diese Bedingungen werden mit zunehmender Länge des Projektionshorizonts immer unrealistischer.

#### **Der Ansatz von Stoto**

Auch Stoto [272] hat 1983 einen Ansatz vorgelegt, mit dem Konfidenzintervalle für zukünftige Bevölkerungsprognosen mittels vergangener Prognosefehler bestimmt werden können. Dabei wird der Fehler der jährlichen Wachstumsrate einer Bevölkerung in zwei Komponenten zerlegt, wobei die eine Komponente den Fehler des Jump-off Jahres und die andere Komponente einen zufälligen Fehlerterm repräsentiert.<sup>10</sup> Beide Fehlerkomponenten weisen eine stabile zeitliche Entwicklung (für entwickelte Länder) auf, sodass Stoto davon ausgeht, dass eine Übertragung von Fehlern vergangener Prognosen auf aktuelle Prognosen überhaupt möglich ist. Die Analyse zahlreicher vergangener Bevölkerungsprognosen der UN ergibt eine Standardabweichung des Fehlers von 0,28. Unter der Annahme einer Normalverteilung kann ein 95 Prozent Konfidenzintervall aus der Summe der prognostizierten Wachstumsrate und der doppelten Standardabweichung des Fehlers für einzelne Projektionsintervalle berechnet werden. Bei der Bestimmung eines 2/3 Konfidenzintervalls wird dann anstatt der doppelten die einfache Standardabweichung des Fehlers verwendet.

---

<sup>9</sup>So schwanken die ermittelten durchschnittlichen Fehlerwerte zwischen  $\pm 0,29$ ,  $\pm 0,48$  und  $\pm 0,60$  bei Bevölkerungen mit einer jährlichen Wachstumsrate von bis zu 1,8 Prozent, zwischen 1,8 und 2,6 Prozent und über 2,6 Prozent.

<sup>10</sup>Die Fehlerkomponente des Jump-off Jahres berücksichtigt den Zeitpunkt der Erstellung einer Bevölkerungsprognose. So kann eine Bevölkerungsprognose einen größeren Prognosefehler erzeugen, wenn sie zu einem Zeitpunkt erstellt worden ist, nach dem unerwartete Veränderungen in der Bevölkerungsentwicklung eingetreten sind. Ein Beispiel in der Geschichte stellen Bevölkerungsprognosen für die USA in den 1940ern dar, die den kurz darauf einsetzenden Babyboom nicht antizipiert und somit eine viel zu geringe Bevölkerungszahl prognostiziert haben.

### **Zusammenfassung**

Die Ansätze zur Fehlerfortschreibung von Keyfitz [136] und Stoto [272] bestimmen einen Unsicherheitsbereich für aktuelle deterministische Bevölkerungsprognosen auf der Basis vergangener Prognosefehler. Dabei unterstellen sie, dass sich weder die Prognosemethoden, noch die Datenqualität (einschließlich der generell verfügbaren Informationen) verbessert hätten, sodass exakt die Fehler vergangener auf aktuelle Bevölkerungsprognosen übertragbar seien. Diese Annahme ist für einen längeren Projektionshorizont nicht haltbar, da die Prognosemethoden fortlaufend weiterentwickelt werden und qualitativ immer bessere Daten verfügbar sind. Daneben verändern sich auch die Faktoren, die eine Bevölkerungsentwicklung maßgeblich beeinflussen. Demzufolge verändern sich die Bedingungen für die Erstellung einer Bevölkerungsprognose über die Zeit derart, dass Fehler vergangener, insbesondere weiter zurückliegender Bevölkerungsprognosen nicht exakt auf aktuelle Bevölkerungsprognosen übertragen werden können. Zumal sich die Unsicherheit einer Bevölkerungsprognose aus der (unbekannten) zukünftigen Entwicklung der demographischen Komponenten bestimmt, sodass weniger Fehler vergangener Bevölkerungsprognosen als vielmehr das jeweilige Spektrum potentieller zukünftiger Entwicklungen als Quelle zur Bestimmung eines Unsicherheitsbereichs herangezogen werden sollte.

### **4.2.3. Extrapolation der Bevölkerung nach Alter und Geschlecht unter Berücksichtigung demographischer Komponenten**

Probabilistische Ansätze zur Bevölkerungsprognose, die die zukünftige Bevölkerung disaggregiert nach Alter und Geschlecht unter Berücksichtigung demographischer Komponenten extrapolieren, sind relativ komplex.

#### **Der Ansatz von Lee und Tuljapurkar**

Lee und Tuljapurkar [153, 159] haben 1994 einen Ansatz zur probabilistischen Prognose einer Bevölkerung nach Alter und Geschlecht auf der Basis der Kohorten-Komponenten Methode (mit einer stochastischen Projektionsmatrix) vorgeschlagen. Zur Fortschreibung einer Bevölkerung (nach der Formel 3.3 auf Seite 38) werden die Mortalität und Fertilität separat mit einem (beliebigen) Modell<sup>11</sup> prognostiziert, sodass die Annahmenwerte (wie

---

<sup>11</sup>So kann z. B. die Mortalität mit dem Lee-Carter Modell [50, 157] (vgl. Abschnitt 5.2.1 ab Seite 84) und die Fertilität nach dem von Lee [155] abgewandelten Lee-Carter Modell zur Prognose der Fertilität (vgl. Abschnitt 6.1.1 ab Seite 117) extrapoliert werden.

auch die Projektionsmatrix) zu jedem Projektionsintervall in einen deterministischen und einen stochastischen Teil zerlegt werden können; der deterministische bzw. stochastische Teil entspricht dabei jeweils dem zu erwartenden Punktprognosewert bzw. der zu erwartenden stochastischen Abweichung (mit dem Mittelwert null). Daneben werden für die Nettomigration lediglich deterministische Annahmenwerte generiert und am Ende eines jeden Projektionsintervalls zur (stochastisch) prognostizierten Bevölkerung addiert. Die Zerlegung der Projektionsmatrix in einen deterministischen und stochastischen Teil ermöglicht auch die Zerlegung des (aus den entsprechenden Matrixprodukten) resultierenden Bevölkerungsvektors in einen deterministischen Term und Terme höherer Ordnung  $n$ , die den Momenten der Ordnung  $n$  der stochastischen Abweichungen aus der Projektionsmatrix entsprechen. So entsprechen beispielsweise die ersten und zweiten (zentralen) Momente dem arithmetischen Mittel und der Varianz der stochastischen Abweichungen vom jeweiligen deterministischen Punktprognosewert, die dann auch zur analytischen Berechnung entsprechender Konfidenzintervalle für diverse Ergebnisgrößen genutzt werden können.

Lee und Tuljapurkar [159] nutzen diesen Ansatz zur probabilistischen Prognose der US-amerikanischen Bevölkerung von 1990 bis 2065, wobei die Zeitreihenmodelle für die Mortalität und Fertilität an vergangene Werte von 1900 bzw. 1909 bis 1989 angepasst worden sind. Entsprechend der getroffenen Annahmen ergibt sich für die totale Bevölkerung im Zieljahr der Prognose ein Wert von 407,5 Millionen, der mit einer Wahrscheinlichkeit von 95 Prozent zwischen 258,7 und 641,8 Millionen schwankt.

#### **Der Ansatz von Hyndman und Booth**

Hyndman und Booth [116] haben 2008 einen probabilistischen Ansatz zur Prognose einer Bevölkerung nach Alter und Geschlecht mit der Kohorten-Komponenten Methode vorgeschlagen. Die Fertilität, Mortalität und Nettomigration sind Zufallsvariablen, deren Annahmenwerte mit einem entsprechenden Unsicherheitsbereich über ein Zeitreihenmodell extrapoliert werden. Dabei werden die alters- und geschlechtsspezifischen Sterberaten und totalen Nettomigrationszahlen wie auch die altersspezifischen Fertilitätsraten der Frauen mit dem leicht veränderten *functional data* Ansatz von Hyndman und Ullah [118] generiert (vgl. Abschnitte 5.2.1 und 6.1.1 auf den Seiten 88 und 118), der jeweils eine Erweiterung der Modelle von Lee und Carter [50, 157] und Lee [155] zur Prognose der Mortalität und Fertilität darstellt (vgl. Abschnitte 5.2.1 und 6.1.1 auf den Seiten 84 und 117).



Um nun eine probabilistische Bevölkerungsprognose mit dem Modell von Hyndman und Booth [116] durchzuführen, wird in jedem Prognosedurchlauf ein Annahmenverlauf mit dem *functional data* Ansatz für die Mortalität, Fertilität und Nettomigration über den gesamten Projektionshorizont zufällig erzeugt. Anschließend wird der entsprechende Ergebnisverlauf der Bevölkerung (schrittweise von Projektionsintervall zu Projektionsintervall) mit der Bevölkerungsbilanzgleichung über die Kohorten-Komponenten Methode berechnet. Dabei werden die Sterbefälle bzw. Geburten aus einer Poissonverteilung, deren Mittelwert aus der zufällig generierten Sterberate bzw. Fertilitätsrate und der jeweiligen durchschnittlichen Risikobevölkerung erstellt wird, zusätzlich zufällig in jedem Projektionsintervall gezogen. Dieses Prozedere wird  $n$  Mal wiederholt, sodass für jede Ergebnisgröße (wie z. B. die totale Bevölkerungszahl) eine Ergebnisverteilung mit  $n$  Ergebnisverläufen resultiert.

#### **Zusammenfassung**

Komplexe Ansätze prognostizieren eine Bevölkerung im Vergleich zu einfachen Ansätzen (vgl. Abschnitt 4.2.1 auf Seite 59) nach Alter und Geschlecht und in Abhängigkeit ihrer demographischen Komponenten (also der Fertilität, Mortalität und Migration).

Dadurch können zwar bei der Prognose der zukünftigen Bevölkerungsentwicklung detaillierte Entwicklungen der Fertilität, Mortalität und Migration (nach Alter und Geschlecht) berücksichtigt werden; jedoch ist die Annahmengenerierung auf in der Vergangenheit beobachtete Werte als einzige (aller verfügbaren) Informationsquelle(n) beschränkt, sobald ausschließlich extrapolative Verfahren dafür eingesetzt werden. Folglich können durch die Limitierung der zu verwendenden Informationsquellen etwaige Trendumkehrungen in der zukünftigen Entwicklung einzelner Modellparameter auch von solchen komplexen Ansätzen nur unvollkommen antizipiert und eingeschlossen werden, sodass auch sie die Prognoseunsicherheit unterschätzen können. Zudem kann die Prognoseunsicherheit durch die Vorgabe komplexer Zeitreihenmodelle zur Annahmengenerierung unterschätzt werden, indem sie bei der Konstruktion der Konfidenzintervalle bzw. der stochastischen Abweichungen nicht alle Fehlerquellen einschließen.<sup>12</sup> Diese Unterschätzung der Prognoseunsicherheit nimmt bei instabilen Bedingungen und demzufolge mit der Länge des Projektionshorizonts zu.

---

<sup>12</sup>Weitere Nachteile, die durch die Verwendung komplexer Zeitreihenmodelle zur Prognose der Mortalität und der Fertilität mit dem (abgewandelten) Lee-Carter Modell entstehen, sind den Abschnitten 5.2.1 und 6.1.1 ab den Seiten 84 und 117 zu entnehmen.

Komplexe probabilistische Ansätze, die eine Bevölkerung nach Alter und Geschlecht unter Berücksichtigung demographischer Komponenten extrapolieren, eignen sich demnach für Kurzzeitprognosen, da sie (fast) ausschließlich die vergangene Entwicklung als Informationsquelle berücksichtigen. Um eine Bevölkerung über einen längeren, möglicherweise instabilen Zeitraum prognostizieren zu können, müssen jedoch weitere Informationsquellen bei einer Bevölkerungsprognose eingeschlossen werden.

#### 4.2.4. Expertenmeinung und Simulation

Neben den extrapolativen Ansätzen, die eine Bevölkerung entweder aggregiert oder disaggregiert auf der Basis der historischen Bevölkerungsentwicklung prognostizieren (und den Ansätzen mit Fehlerfortschreibung), gibt es Ansätze mit Expertenmeinung, die darüber hinaus auch noch weitere Informationen bei einer Bevölkerungsprognose berücksichtigen (und korrigierend einsetzen).

##### Der Ansatz von Pflaumer

Pflaumer [215] hat 1988 einen Ansatz zur probabilistischen Bevölkerungsprognose entwickelt, der auf der simulativen Ausführung der Kohorten-Komponenten Methode mit Zufallsvariablen und Expertenmeinung basiert. So werden die TFR, die Lebenserwartung bei Geburt, die totale Nettomigration wie auch die maximale absolute Abweichung zwischen zwei aufeinanderfolgenden TFR-Werten als Zufallsvariablen behandelt. Mit Expertenmeinung wird für jede dieser Zufallsvariablen ein zukünftiger Wertebereich mit einem Median, einem Minimum und einem Maximum subjektiv bestimmt. Darüber hinaus wird den Werten der Zufallsvariablen TFR, Lebenserwartung bei Geburt und totale Nettomigration eine entsprechende Altersverteilung zugeordnet.<sup>13</sup> Um nun die Bevölkerung nach dem Alter in einem Projektionsintervall zu prognostizieren, wird zufällig ein Wert für jede Zufallsvariable gezogen. Dabei hat jeder Wert innerhalb eines Wertebereichs dieselbe Wahrscheinlichkeit, zufällig gezogen zu werden. Die Bevölkerungszahl in einem Projektionsintervall ergibt sich dann aus der Bevölkerung des vorherigen Projektionsintervalls,<sup>14</sup> der die im Projektionsintervall stattfindenden Geburten und Nettozuzüge hinzugerechnet und von der die im Projektionsintervall stattfindenden Sterbefälle abgezogen werden. Dieses Prozedere wird

---

<sup>13</sup>Der TFR werden altersspezifische Fertilitätsraten, der Lebenserwartung bei Geburt werden altersspezifische Überlebenswahrscheinlichkeiten und der totalen Nettomigration werden altersspezifische Nettomigrationszahlen zugeordnet.

<sup>14</sup>Im ersten Projektionsintervall wird ersatzweise die Jump-off Bevölkerung eingesetzt.

schrittweise für jedes Projektionsintervall bis zum Zieljahr der Prognose durchgeführt, sodass ein Ergebnisverlauf für die zukünftige Bevölkerungszahl resultiert. Um nun aber eine Ergebnisverteilung für die zukünftige Bevölkerungsentwicklung zu erhalten, werden  $n$  Ergebnisverläufe der zukünftigen Bevölkerungszahl berechnet, sodass Wahrscheinlichkeitsaussagen über die Größe einer zukünftigen Bevölkerung getroffen werden können. Pflaumer [215] verwendet diesen Ansatz zur Prognose der US-amerikanischen Bevölkerung von 1987 bis 2082 in Projektionsintervallen der Länge von fünf Jahren. Ausgangspunkt ist die Jump-off Bevölkerung aus dem Jahr 1982. Nach der Berechnung von 1.000 Ergebnisverläufen wird demnach die US-amerikanische Bevölkerung im Zieljahr 2082 mit einer Wahrscheinlichkeit von 90 Prozent zwischen 254,6 Millionen und 354,8 Millionen Personen (nach diesem Ansatz und mit den getroffenen Annahmen) liegen.

#### **Der Ansatz von Lutz, Sanderson und Scherbov**

Lutz et al. [173] haben 1996 einen umfassenden Ansatz zur probabilistischen Bevölkerungsprognose vorgelegt, der auf der simulativen Ausführung der Kohorten-Komponenten Methode mit Zufallsvariablen und Expertenmeinung basiert. Für die Modellparameter TFR, Lebenserwartung bei Geburt und Nettomigration wird jeweils ein Bereich potentieller Annahmenwerte mit einem Minimum, einem Maximum und einem Mittelwert von Experten für ein Projektionsintervall einer Prognose abgesteckt. Anschließend wird dieser Wertebereich von Experten in eine Wahrscheinlichkeitsverteilung<sup>15</sup> übertragen, wobei auch angegeben werden muss, wie groß der Anteil des anzunehmenden Wertebereichs an der gesamten Wahrscheinlichkeitsverteilung ist. Deckt der Wertebereich weniger als 100 Prozent ab, ist die Streuung um den Mittelwert mit der anzunehmenden Wahrscheinlichkeitsverteilung größer als vom zuerst abgesteckten Wertebereich. Bei der Durchführung eines Prognosedurchlaufs wird für jeden Modellparameter zufällig ein Annahmenwert (in Abhängigkeit seiner Eintrittswahrscheinlichkeit) aus der jeweiligen Wahrscheinlichkeitsverteilung gezogen. Daraufhin kann für jeden Modellparameter ein Annahmenverlauf mit einer stückweisen linearen Interpolation vom Startwert (aus dem Startjahr) zum zufällig zu ziehenden Wert (aus der Wahrscheinlichkeitsverteilung) über den gesamten Projektionshorizont generiert werden. Aus diesen Annahmenwerten werden dann diverse Ergebnisgrößen mit der Kohorten-Komponenten Methode für jedes Projektionsintervall berechnet. Durch die Berechnung  $n$  unabhängiger Prognosedurchläufe entstehen für jede Ergebnisgröße  $n$  Ergebnisverläufe, die

---

<sup>15</sup>In praktischen Anwendungen wird häufig die Normalverteilung als Wahrscheinlichkeitsverteilung verwendet.

einer Ergebnisverteilung mit Eintrittswahrscheinlichkeiten entsprechen.

Lutz et al. [176] nutzen 1998 diesen Ansatz zur Prognose der Bevölkerung in Deutschland nach Einzelalter von 1995 bis 2050.<sup>16</sup> Dazu wird von Experten für jeden Modellparameter eine Normalverteilung mit potentiellen Annahmenwerten für das Projektionsintervall 2030 konstruiert. Bei der Fertilität wird ein Wertebereich für die TFR festgelegt, der zwischen den beiden Grenzwerten 1 und 1,9 um den Mittelwert von 1,45 schwankt. Dieser Wertebereich soll 90 Prozent der Normalverteilung im Jahr 2030 abdecken. Zur Erstellung eines Annahmenverlaufs für die Fertilität wird zwischen dem Startwert und dem zufällig zu ziehenden Wert aus der Normalverteilung stückweise linear interpoliert, wobei in den ersten Projektionsintervallen ein stärkerer linearer Anstieg angenommen wird, der sich mit zunehmender Länge des Projektionshorizonts verringert.<sup>17</sup> Bei der Mortalität wird von den Experten ein Anstieg zwischen 0 und 3 Jahren für die Lebenserwartung bei Geburt pro Dekade angenommen. Dieser Wertebereich, der um einen Mittelwert von 1,5 Jahren schwankt, soll ebenfalls 90 Prozent der Normalverteilung im Jahr 2030 abdecken. Bei der Migration wird von den Experten ein Wanderungssaldo angenommen, der zwischen 100.000 und 300.000 Personen um einen Mittelwert von 200.000 Personen schwankt. Da die Prognoseunsicherheit bei der Migration besonders hoch ist, gehen die Experten davon aus, dass dieser Wertebereich nur 67 Prozent der Normalverteilung im Jahr 2030 abdeckt. Nach 1.000 unabhängigen Prognosedurchläufen schwankt die Bevölkerungszahl in Deutschland im Jahr 2050 mit einer Wahrscheinlichkeit von 95 Prozent zwischen 51,25 Millionen und 93,72 Millionen Personen um einen Mittelwert von 71,76 Millionen Personen.

Lutz et al. [174, 244] haben 2001 ihren ersten Ansatz zur probabilistischen Bevölkerungsprognose mit Expertenmeinung mit dem Ziel weiterentwickelt, die Varianz der anzunehmenden Modellparameterwerte zu erhöhen bzw. nicht weiter zu unterschätzen.<sup>18</sup> Dazu haben sie die modell-basierte Methode zur Generierung der Annahmenwerte für die einzelnen Modellparameter verändert. Anstatt die Annahmenwerte von einem Startwert zum

---

<sup>16</sup>Lutz et al. [173, 175] nutzen diesen probabilistischen Ansatz zur Bevölkerungsprognose aber nicht nur für Deutschland, sondern z. B. auch für 13 Weltregionen wie auch für Österreich.

<sup>17</sup>Zwischen dem zufällig zu ziehenden Wert für das Jahr 2030 und dem Zieljahr der Prognose wird eine Extrapolation verwendet.

<sup>18</sup>So hat z. B. Lee [156] den Ansatz von Lutz et al. [173] dahingehend kritisiert, dass durch die Verwendung einer stückweisen linearen Interpolation die Varianz bei der Generierung der Annahmenverläufe unterschätzt werde.

zufällig zu ziehenden Zielwert stückweise linear zu interpolieren, werden sie nun über einen subjektiv zu bestimmenden Mittelwert und eine mit einem stochastischen Zeitreihenprozess zufällig zu generierende Abweichung, deren potentiell Ausmaß mit Expertenmeinung vorzugeben ist, in jedem Projektionsintervall neu bestimmt. Die simulative Ausführung der Bevölkerungsprognose bleibt erhalten, sodass für jede Ergebnisgröße eine Ergebnisverteilung aus  $n$  Prognosedurchläufen resultiert.

Generell wird für die Modellparameter TFR, Veränderung der Lebenserwartung bei Geburt in Jahren und Nettomigration, jeweils ein Mittelwert pro Projektionsintervall mit Expertenmeinung vorher bestimmt. Damit die Annahmenwerte für jeden Modellparameter,  $v_t$ , um den jeweils vorzugebenden Mittelwert,  $\bar{v}_t$ , schwanken können, müssen zudem in jedem Projektionsintervall zufällige Abweichungen,  $\varepsilon_t$ ,<sup>19</sup> generiert werden (vgl. Formel 4.1).

$$v_t = \bar{v}_t + \varepsilon_t \quad (4.1)$$

Diese zufälligen Abweichungen werden mit einem *Moving Average* ( $MA(q)$ )-Prozess erzeugt, bei dem mit Expertenmeinung die Breite der Fehlerverteilung über  $\sigma(\varepsilon)$ , die Streuung um den vorzugebenden Mittelwert über  $\sigma(u)$ <sup>20</sup> und die Anzahl der zu berücksichtigenden vorherigen Fehlerwerte,  $q$ , bestimmt werden können. Je höher die  $\sigma(\varepsilon)$ , desto größer sind die maximalen Abweichungen vom vorzugebenden Mittelwert und je kleiner die  $\sigma(u)$ , desto dichter streuen die Abweichungen um den vorzugebenden Mittelwert (vgl. Formel 4.2).

$$\varepsilon_t = \sum_{i=0}^q \frac{\sigma(\varepsilon)}{\sqrt{q+1}} u_{t-i} \quad (4.2)$$

Mit dem  $MA(q)$ -Prozess können Autokorrelationen der Abweichungen über den Projektionshorizont modelliert werden, da bei der Bestimmung der Abweichung in einem Projektionsintervall  $t$  auch die zufälligen Abweichungen  $q$  vorhergehender Projektionsintervalle berücksichtigt werden. Zudem können durch die Generierung korrelierender Zufallszahlen für jedes Projektionsintervall Korrelationen zwischen verschiedenen Modellparametern (wie z. B. zwischen der TFR und der Lebenserwartung bei Geburt) innerhalb einer Region oder auch für einen Modellparameter (wie z. B. die TFR) zwischen verschiedenen

---

<sup>19</sup>Die  $\varepsilon_t$  sind normalverteilte Zufallsvariablen mit einem Mittelwert von *null* und einer Standardabweichung  $\sigma(\varepsilon)$ .

<sup>20</sup>Die  $u_{t-i}$  sind unabhängige und standardnormalverteilte Zufallsvariablen mit einem Mittelwert *null* und einer Standardabweichung  $\sigma(u)$ .

Regionen unterstellt werden.

Lutz et al. [174] nutzen diesen verbesserten Ansatz zur Prognose der Bevölkerungszahl nach Einzelalter für die Welt wie auch für 13 Weltregionen vom Startjahr 2000 bis zum Zieljahr 2100 in Projektionsintervallen der Länge von 5 Jahren. Dabei basieren ihre Annahmen für die Fertilität, Mortalität und Migration hauptsächlich auf früheren Bevölkerungsprognosen mit dem alten Prognosemodell [173], deren Annahmen überarbeitet bzw. aktualisiert worden sind. So werden bei der Fertilität zunächst die Mittelwerte für die zwei Projektionsintervalle 2025-29 und 2080-84 mit Expertenmeinung bestimmt, um anschließend die Mittelwerte der übrigen Projektionsintervalle mit einer Interpolation (zwischen den jeweils gegebenen Mittelwerten) festzulegen. Danach wird die Varianz der Fertilität in Abhängigkeit von der Höhe der TFR festgelegt, wobei z.B. die Abweichung einer TFR von über 3 mit einer Wahrscheinlichkeit von 80 Prozent zwischen  $+/-1$  liegen soll. Bei der Mortalität gehen Lutz et al. [174] davon aus, dass die Lebenserwartung bei Geburt alle 10 Jahre im Durchschnitt um 2 Jahre steigt, wobei die Abweichung mit einer Wahrscheinlichkeit von 80 Prozent  $+/-2$  Jahre betragen soll. Die Annahmen zur Migration stehen in einem Zufallsvektor, der typische interregionale Nettomigrationszahlen der (erst kürzlich) vergangenen Jahre enthält. Lutz et al. [174] nutzen generell einen MA(31)-Prozess zur Modellierung der Abweichungen von den jeweils anzunehmenden Mittelwerten bei allen Modellparametern, sodass jeweils die Abweichungen der 31 vorangegangenen Projektionsintervalle bei der Bestimmung der aktuellen Abweichung berücksichtigt werden. Darüber hinaus gehen Lutz et al. [174] von interregionalen Korrelationen bei der Fertilität in der Höhe von 0,7 und bei der Mortalität in der Höhe von 0,9 aufgrund der zunehmenden Globalisierung aus. Dabei begründen sie die höhere Korrelation bei der Mortalität im Vergleich zur Fertilität damit, dass Verhaltensänderungen (bezüglich der Fertilität) langsamer vonstatten gehen als der Gebrauch neuer medizinischer Technologien oder die Ausbreitung von Gesundheitsrisiken. Weiterhin nehmen Lutz et al. [174] eine perfekte Korrelation bei der Lebenserwartung bei Geburt von Männern und Frauen an. Nach der Berechnung von insgesamt 2.000 Prognosedurchläufen schwankt die Weltbevölkerungszahl (nach den getroffenen Annahmen) mit 80 Prozent Wahrscheinlichkeit zwischen 5,577 Milliarden und 12,123 Milliarden Personen um einen Median von 8,414 Milliarden Personen im Zieljahr 2100.

Ebenso wie Lutz et al. [174] nutzen auch Scherbov et al. [246] 2008 diesen überarbeiteten Ansatz zur probabilistischen Bevölkerungsprognose, der als *expert argument-based approach*

to probabilistic forecasting bezeichnet wird. Dabei wird die Bevölkerungsentwicklung von 27 EU-Mitgliedsländern vom Jump-off Jahr 2006 bis zum Zieljahr 2050 nach Einzelalter und Geschlecht prognostiziert. Die Annahmen zur Fertilität, Mortalität und Migration basieren auf den Annahmen der Bevölkerungsprognosen von Eurostat von 2005 bis 2051 nach der Szenario-Technik [148]. Zur Bestimmung der jeweiligen Mittelwerte,  $\bar{v}_t$ , verwenden Scherbov et al. [246] das Level aus dem Jump-off Jahr für die TFR und das jeweilige *baseline* Szenario für die Lebenserwartung bei Geburt wie auch für die Nettomigration. Zur Bestimmung der Abweichungen von jedem Mittelwert muss jeweils ein Unsicherheitsbereich bestimmt werden. Die hohen und niedrigen Annahmenverläufe der Eurostat-Prognose werden dafür in einen symmetrischen, normalverteilten Unsicherheitsbereich (um den jeweiligen Mittelwert) übertragen, wobei mit Expertenmeinung zusätzlich festgelegt werden muss, welchen Anteil der Bereich zwischen den Grenzverläufen am gesamten, anzunehmenden, normalverteilten Unsicherheitsbereich tatsächlich abdeckt. So werden bei den Annahmen zur Fertilität 70 Prozent, bei den Annahmen zur Mortalität 60 Prozent und bei den Annahmen zur Migration nur noch 50 Prozent Abdeckung angenommen. Diese Angaben können über die Standardabweichung der Abweichungen vom Mittelwert,  $\sigma(\varepsilon)$ , eingegeben werden, wobei die Abweichungen,  $\varepsilon_t$ , generell mit einem MA(41)-Prozess generiert werden. Nach der Berechnung von jeweils 1.000 unabhängigen Prognosedurchläufen ergibt sich z. B. für Deutschland im Zieljahr 2050 eine zentrale Bevölkerung von 73,44 Millionen Personen, die mit einer Wahrscheinlichkeit von 80 Prozent zwischen 61,08 Millionen und 86,09 Millionen Personen schwankt.

#### **Zusammenfassung**

Die probabilistischen Bevölkerungsprognoseansätze von Pflaumer [215] und Lutz et al. [173, 174] nutzen die Kohorten-Komponenten Methode, um eine Bevölkerung nach Alter und Geschlecht auf der Basis demographischer Ereignisse zu prognostizieren. Dabei werden die Modellparameter zur Fertilität, Mortalität und Migration als Zufallsvariablen behandelt, deren Unsicherheitsbereich von Experten bestimmt wird. Die subjektive Bestimmung des Unsicherheitsbereichs eines Modellparameters ermöglicht eigentlich die uneingeschränkte Modellierung aller potentiell möglichen Annahmenverläufe. Allerdings wird das bei diesen Ansätzen für jeden Modellparameter zu berücksichtigende potentielle Entwicklungsspektrum durch die (grobe) Vorgabe des Verlaufs und des Niveaus über eine modell-basierte Methode oder/und durch die Zuordnung einer Wahrscheinlichkeitsverteilung eingeschränkt. So geht Pflaumer [215] von einer Gleichverteilung aus, bei der alle Annahmenwerte (und -verläufe) die gleiche Eintrittswahrscheinlichkeit erhalten. Im Gegensatz dazu gehen Lutz et al.

[173, 174] bei praktischen Anwendungen ihres ersten und zweiten Ansatzes häufig von einer Normalverteilung aus, sodass es jeweils nur einen wahrscheinlichsten Annahmenwert (und -verlauf) gibt, von dem symmetrisch nach oben und unten abweichende Annahmenwerte mit abnehmender Wahrscheinlichkeit generiert werden können. Der überarbeitete Ansatz von Lutz et al. [174] lässt zwar zufällige Schwankungen in jedem Projektionsintervall zu, allerdings nur um einen jeweils subjektiv fest vorzugebenden Mittelwert. Gibt es nur einen potentiellen Annahmenverlauf mit einer höchsten Eintrittswahrscheinlichkeit pro Modellparameter, können diese Ansätze dessen zukünftiges Entwicklungsspektrum angemessen abdecken. Sofern es jedoch bei einer Prognose für einen Modellparameter mehrere, auch stark voneinander abweichende potentielle Annahmenverläufe (mit ähnlich hohen Eintrittswahrscheinlichkeiten) gibt, die sich sowohl im Niveau als auch in der Verlaufsstruktur deutlich voneinander unterscheiden, kann dessen zukünftiges Entwicklungsspektrum von diesen Ansätzen nicht mehr angemessen eingefangen werden. Annahmenverläufe eines Modellparameters, die sich stark im Niveau, im Verlauf wie auch in der zu erwartenden Eintrittswahrscheinlichkeit unterscheiden, treten bei besonders unsicheren Prognosen (wie beispielsweise bei der Prognose der Migration (vgl. Kapitel 7)) auf und sollten variabel in einem Bevölkerungsprognosemodell eingefangen und berücksichtigt werden können, um die Prognosegenauigkeit zu steigern (vgl. Abschnitt 10.2.1 ab Seite 200).

Darüber hinaus modellieren alle drei Ansätze den Modellparameter der Nettomigration relativ oberflächlich, sodass die bevölkerungsdynamischen Effekte tatsächlich auftretender Zu- und Fortzüge, die sich sowohl in der Alters- und Geschlechtsstruktur als auch im demographischen Verhalten (untereinander und von den Einheimischen) unterscheiden können, vereinfachend ignorieren, wodurch die Prognoseergebnisse erheblich verzerrt werden können.

Weiterhin ist bei dem probabilistischen Bevölkerungsprognoseansatz von Lutz et al. [173, 174] der Rückschluss von den zufällig zu generierenden Annahmenwerten der Zufallsvariablen TFR, Lebenserwartung bei Geburt und Nettomigration auf die jeweilige Altersstruktur der tatsächlich in der Kohorten-Komponenten Methode zu verwendenden Modellparameterwerte unklar. In der Anwendung dieses Ansatzes von Mamolo et al. [181] wird von einer konstanten Altersstruktur (des letzten verfügbaren Jahres oder auch eines Durchschnittswerts mehrerer Länder (und Jahre)) gesprochen. Dies wäre jedoch eine starke Vereinfachung, die ebenfalls zu verzerrten Prognoseergebnissen führen kann.



Generell sind die Ansätze von Pflaumer [215] und Lutz et al. [173, 174] vielversprechende komplexe Bevölkerungsprognosemodelle, die die zukünftige Bevölkerung auf der Makroebene nach der Zahl und Struktur nicht nur mit quantitativen Methoden wie z. B. der Extrapolation, sondern auch unter Einsatz weiterer Informationsquellen wie der Expertenmeinung fortschreiben. Zudem erfolgt die Berechnung einer Bevölkerungsprognose mit diesen Ansätzen simulativ, sodass mehrere (meist) verschiedene Prognosedurchläufe mit der jeweils modell-basierten Methode zur Annahmengenerierung zufällig erzeugt werden, woraus letztlich für jede Ergebnisgröße eine Ergebnisverteilung resultiert. Die eingebaute Zufälligkeit wie auch die Nutzung von Expertenmeinung stellen herausragende Merkmale dieser drei Ansätze dar, da sie eine flexible Annahmengenerierung ermöglichen. Allerdings reicht diese (z. T. eingeschränkte) Flexibilität genau dann nicht aus, wenn aufgrund einer hohen Prognoseunsicherheit eine Annahmenverteilung mit heterogenen und variabel gewichteten Annahmenverläufen für jeden Modellparameter erstellt werden soll. Zudem können durch die oberflächliche Ausarbeitung der Migration wesentliche bevölkerungsdynamische Effekte mit diesen drei Ansätzen nicht eingefangen werden. Demzufolge ist die eingeschlagene Richtung bei der Bevölkerungsprognose mit diesen Ansätzen wegweisend, aber auch noch erweiterbar, um die Prognoseunsicherheit umfassender und gezielter einfangen zu können.

##### 4.2.5. Verbindung verschiedener Methoden

Eine weitere Gruppe von probabilistischen Ansätzen zur Bevölkerungsprognose nutzt nicht nur eine, sondern mehrere Methoden zur Generierung der Annahmen, um die Prognoseunsicherheit umfassend einfangen und quantifizieren zu können.

##### Der Ansatz von Alho und Spencer

Alho und Spencer [4, 6, 8, 9, 10] haben Ende der 1990er ein probabilistisches Bevölkerungsprognosemodell vorgeschlagen, das auf der simulativen Ausführung der Kohorten-Komponenten Methode mit Zufallsvariablen basiert. Zur Generierung der Annahmen können neben der Expertenmeinung z. B. auch empirische Fehler vergangener Prognosen (vgl. Abschnitt 8.1 ab Seite 147) herangezogen werden.

In dem Modell von Alho und Spencer werden die alters- und geschlechtsspezifischen Fertilitäts- und Mortalitätsraten wie auch die totalen Nettomigrationszahlen (nach Alter und Geschlecht) als Zufallsvariablen behandelt. Für jede dieser Zufallsvariablen müssen in

einem ersten Schritt Punktprognosewerte für jedes Projektionsintervall extern (mit einer beliebigen Methode) generiert werden.

Anschließend werden die tatsächlichen zukünftigen Modellparameterwerte im Alter  $x$  im Projektionsintervall  $t$ ,  $R_{x,t}$ , aus dem jeweiligen Punktprognosewert,  $F_{x,t}$ , und dem entsprechenden im Modell zu simulierenden Fehler,  $E_{x,t}$ , (zufällig) generiert (vgl. Formeln 4.3 und 4.4).

$$R_{x,t} = F_{x,t} \cdot e^{E_{x,t}} \quad (4.3)$$

$$\ln R_{x,t} = \ln F_{x,t} + E_{x,t} \quad (4.4)$$

Der Fehler  $E_{x,t}$  eines zu prognostizierenden Modellparameterwertes im Alter  $x$  im Projektionsintervall  $t$  setzt sich aus der Summe der einzelnen Fehlerterme  $\varepsilon_{x,t}$  im Alter  $x$  vom Startjahr bis zum jeweiligen Projektionsintervall  $t$  zusammen (vgl. Formel 4.5). Der Erwartungswert der Fehlerterme  $\varepsilon_{x,t}$  beträgt null.

$$E_{x,t} = \varepsilon_{x,1} + \varepsilon_{x,2} + \dots + \varepsilon_{x,t} \quad (4.5)$$

Das so genannte *Scaled Model for Error* wird zur Simulation der Prognosefehler verwendet. Jeder Fehlerterm  $\varepsilon_{x,t}$  besteht aus zwei zufälligen Komponenten, deren Summe über einen deterministischen Skalenwert hoch- oder herunterskaliert wird (vgl. Formel 4.6). Die beiden zufälligen Komponenten bilden die Autokorrelation ( $\delta_{x,t}$ ) und die Kreuzkorrelation ( $n_x$ ) ab, deren Varianz über den Wert von  $\kappa_x$  im Vorhinein festgelegt wird.

$$\varepsilon_{x,t} = S_{x,t} \cdot (n_x + \delta_{x,t}) \quad (4.6)$$

Die beiden zufälligen Fehlerkomponenten sind weder von der Zeit noch voneinander abhängig; zudem sind sie um den Mittelwert null normalverteilt, sodass ihre Schwankungsbreite ausschließlich von der Streuung der Normalverteilung über den Wert von  $\kappa_x$  bestimmt wird: Je größer  $\kappa_x$  ist, desto mehr Gewicht erhält die Fehlerkomponente der Kreuzkorrelation und umgekehrt (vgl. Formeln 4.7 und 4.8).

$$n_x \sim N(0, \kappa_x) \quad (4.7)$$

$$\delta_{x,t} \sim N(0, 1 - \kappa_x) \quad (4.8)$$

Der Modellparameter  $\kappa_x$  gibt für jedes Alter  $x$  das Ausmaß der Korrelation zwischen den Fehlertermen  $\varepsilon_{x,t}$  über den Projektionshorizont fest vor. Der Wert von  $\kappa_x$  liegt zwischen 0 und 1. Ein  $\kappa_x$  von 0 bzw. 1 bedeutet, dass es keine bzw. eine perfekte Korrelation zwischen den  $\varepsilon_{x,t}$  gibt, wohingegen ein  $\kappa_x$  zwischen 0 und 1 eine systematische Fehlerentwicklung über den Projektionshorizont bedeutet (vgl. Formel 4.9).

$$\kappa_x = \text{Corr}(\varepsilon_{x,t}, \varepsilon_{x,t+n}) \quad n \neq 0 \quad (4.9)$$

Der deterministische Skalenwert  $S_{x,t}$  in Abhängigkeit vom Alter  $x$  und dem Projektionsintervall  $t$  ist die fest vorgegebene Standardabweichung der Fehlerterme  $\varepsilon_{x,t}$  (vgl. Formel 4.10). Diese Standardabweichung soll mit der Zeit  $t$  zunehmen. Die generelle Abhängigkeit der  $S_{x,t}$  von der Zeit  $t$  erlaubt eine flexible und variable Modellierung der Varianz der Fehler mit der Zeit.

$$S_{x,t}^2 = \text{Var}(\varepsilon_{x,t}) \quad (4.10)$$

Darüber hinaus können für jeden Modellparameter Annahmen zur Korrelation der Annahmenwerte zwischen den Geschlechtern getroffen werden.

Im Vergleich zur Mortalität und Fertilität werden die zu simulierenden Fehler bei der Nettomigration nach dem Geschlecht  $j$  und dem Projektionsintervall  $t$ ,  $\varepsilon_{j,t}$ , zufällig erzeugt, wobei eine konstante Altersstruktur,  $g_{j,x}$  für die Bruttomigration (, die hier der Summe aus Zu- und Fortzügen entspricht,) berücksichtigt wird (vgl. Formel 4.11). Mangelnde Datenqualität stellen die Gründe für eine derart einfache Modellierung der Migration in diesem Bevölkerungsprognosemodell dar.

$$Y_{j,x,t} = S_{j,t} \cdot g_{j,x} \cdot (n_j + \delta_{j,t}) \quad (4.11)$$

Es gibt verschiedene Möglichkeiten, um die deterministischen Annahmenwerte der  $S_{x,t}$  und  $\kappa_x$  für die Fehlerterme der Mortalität, Fertilität und Nettomigration zu bestimmen. Eine Möglichkeit besteht darin, empirische Fehler vergangener oder/und naiver Prognosen (vgl. Abschnitte 4.2.2 und 8.1 ab den Seiten 61 und 147) für Projektionshorizonte verschiedener Länge zu analysieren. Das Ausmaß, die Streuung und die Korrelation dieser empirischen Prognosefehler dienen dann als Hinweis für Fehler gegenwärtiger Prognosen. Eine weitere Möglichkeit zur Bestimmung der Werte für  $S_{x,t}$  und  $\kappa_x$  besteht neben der Analyse empirischer Prognosefehler in der modell-basierten Fehlerschätzung mittels mathematisch-statistischer

Methoden wie z. B. einem stochastischen Zeitreihenmodell. Expertenmeinung kann zudem genutzt werden, um die mit einer Methode ermittelten Werte für die  $S_{x,t}$  und  $\kappa_x$  zu korrigieren bzw. an bestehende Theorien und Überzeugungen anzupassen [4, 7, 9].

Das *Scaled Model for Error* ist in dem Computerprogramm *Program for Error Propagation* (PEP) implementiert [258]. PEP ist im Rahmen des Projektes *Changing Population of Europe: Uncertain Future*, kurz *Uncertain Population of Europe* (UPE), zur Erstellung probabilistischer Bevölkerungsprognosen für 18 Länder der *European Economic Area* (EEA) mit der Schweiz, aber ohne Liechtenstein verwendet worden. Bei der probabilistischen Bevölkerungsprognose für jedes dieser Länder werden 3.000 Simulationsläufe durchgeführt. In jedem Simulationslauf wird dabei für jede Zufallsvariable ein Wert auf der Basis der Punktprognose und der Unsicherheitsparameter ( $n_x$  und  $\delta_{x,t}$ ) zufällig gezogen. Diese stochastisch variierenden Modellparameterwerte für die Fertilität, Mortalität und die Migration werden dann unter Anwendung der Kohorten-Komponenten Methode zu diversen Ergebnisgrößen wie z. B. dem Bevölkerungsstand in einem Projektionsintervall  $t$  zusammengefasst. Die Ergebnisverteilung bzw. die *predictive distribution* für die jeweilige Ergebnisgröße resultiert aus den korrespondierenden 3.000 Ergebniswerten aller Simulationsläufe. Ein *predictive interval* mit der Eintrittswahrscheinlichkeit von  $(1 - \alpha)100$  Prozent kann für jede Ergebnisgröße so bestimmt werden, dass jeweils  $\frac{\alpha}{2}100$  Prozent der aufsteigend sortierten Ergebniswerte über und unter der oberen und unteren Grenze liegen. Im Anschluss können diese *initial guesses* mit empirischer Fehleranalyse und/oder Expertenmeinung korrigiert werden [4, 7, 261].

Das probabilistische Bevölkerungsprognosemodell von Alho und Spencer hebt sich durch die relativ flexible Kombination verschiedener statistischer und wertender Methoden von allen bisher veröffentlichten probabilistischen Bevölkerungsprognosemodellen ab. Die stochastische Variabilität wird durch einen zufälligen Fehlerprozess um fest vorgegebene Punktprognosewerte erzielt. Diese Fehler schwanken in Abhängigkeit von Modellparametern, deren Werte auf der Analyse empirischer Prognosefehler, (meist) historischer Zeitreihenanalyse und Expertenmeinung basieren. Demnach ist eine außerordentlich hohe Vielfalt der Informationsquellen gegeben. Jedoch wird die Expertenmeinung tatsächlich (häufig) nur marginal zur Korrektur bzw. Anpassung der vornehmlich aus statistischen Methoden resultierenden Modellparameterwerte verwendet. Zudem wird für jeden Modellparameter der Fertilität, Mortalität und Migration jeweils ein fester Punktprognosewert pro Projektionsintervall vorgegeben, sodass

das Niveau, der Verlauf wie auch die Eintrittswahrscheinlichkeit der potentiellen Annahmenverläufe trotz der zufälligen Schwankungen relativ fest vorgegeben sind. So können auch bei diesem Ansatz nicht sich im Niveau und im Verlauf stark unterscheidende Annahmenverläufe variabel gewichtet werden, um eine heterogene Annahmenverteilung bei einer besonders unsicheren Bevölkerungsprognose zu erzeugen (vgl. Abschnitt 10.2.1 ab Seite 200). Außerdem wird die Migration relativ oberflächlich modelliert, da lediglich die totale Nettomigration mit einer festen Altersverteilung für Männer und Frauen prognostiziert wird. Dadurch können die bevölkerungsdynamischen Effekte tatsächlich auftretender Zu- und Fortzüge, die sich sowohl in der Alters- und Geschlechtsstruktur als auch im demographischen Verhalten (untereinander und von den Einheimischen) unterscheiden können, nur ungenügend berücksichtigt werden, sodass die Prognoseergebnisse erheblich verzerrt werden können. Alders et al. [4] räumen zwar diesen Sachverhalt ein und begründen ihn mit der vorliegenden schlechten Datenqualität. Nichtsdestotrotz können dadurch die Prognoseergebnisse (insbesondere für Bevölkerungen, für die detaillierte Daten zur Bruttomigration verfügbar sind) verzerrt werden.

### 4.3. Schlussfolgerung — Anforderungen an das PPPM

Es gibt eine Vielzahl bereits bestehender probabilistischer Ansätze zur Bevölkerungsprognose auf der Makroebene. Das herausstechende Merkmal aller probabilistischen Bevölkerungsprognoseansätze ist die Erzeugung einer Ergebnisverteilung für die jeweiligen Ergebnisgrößen, sodass die Prognoseunsicherheit nicht nur eingefangen, sondern sogar quantifiziert werden kann. Allerdings unterscheiden sich die verschiedenen probabilistischen Ansätze in der Berücksichtigung häufig auftretender Fehler- und Unsicherheitsquellen bzw. in zahlreichen Arbeitsschritten eines Prognoseprozesses (vgl. Abschnitt 2.2 ab Seite 18).

Aus der Analyse der vielfältigen Fehler- und Unsicherheitsquellen einer Bevölkerungsprognose (vgl. Abschnitt 4.1 ab Seite 52) im Allgemeinen und der Vor- und Nachteile der bestehenden probabilistischen Ansätze zur Bevölkerungsprognose (vgl. Abschnitt 4.2 ab Seite 57) im Besonderen können wesentliche Anforderungen an das in dieser Arbeit zu entwickelnde probabilistische Bevölkerungsprognosemodell PPPM abgeleitet werden.

Im Wesentlichen sollte mit dem PPPM eine Bevölkerung genau und sicher prognostiziert werden können. Dabei sollten die Modellparameter derart detailliert ausgestaltet sein, dass z. B. im Gegensatz zu herkömmlichen Prognosemodellen die bevölkerungsdynamischen

Effekte der Zu- und Fortzüge, die sich in der Alters- und Geschlechtsstruktur wie auch im demographischen Verhalten (untereinander und von den Einheimischen) unterscheiden, detailliert berücksichtigt werden können. Darüber hinaus sollten möglichst alle potentiellen zukünftigen Entwicklungen eines jeden Modellparameters flexibel berücksichtigt werden können, sodass die zu erzeugenden Annahmenverläufe einerseits im Niveau und Verlauf (wenn nötig) deutlich voneinander abweichen und andererseits auch variabel gewichtet werden können.

Um eine Bevölkerung also möglichst genau und sicher, d. h. unter Einschluss wesentlicher Unsicherheitsquellen prognostizieren zu können, sollten jeweils alle potentiellen zukünftigen Entwicklungen aller Modellparameter flexibel in einer Prognose mit dem PPPM eingefangen werden können. Demzufolge sollte die Anzahl der einzuschließenden Annahmen für einen Modellparameter im PPPM uneingeschränkt sein. Darüber hinaus kann es erforderlich sein, sehr unterschiedliche Annahmenverläufe für einen Modellparameter in einer Prognose zu berücksichtigen, insbesondere wenn dessen zukünftige Entwicklung besonders unsicher erscheint. So können z. B. die verfügbaren Informationsquellen in ihrer Vorhersage über die zukünftige Entwicklung eines Modellparameters deutlich voneinander abweichen, sodass sich die Annahmenverläufe im Niveau, im Verlauf wie auch in ihrer zu erwartenden Eintrittswahrscheinlichkeit voneinander unterscheiden. Kann ein Prognostiker nun aber keinen dieser unterschiedlichen Annahmenverläufe mit Sicherheit ausschließen, sollte er alle flexibel berücksichtigen können. Dazu sollten jedoch weder eine (oder mehrere) modellbasierte Methode(n) zur Generierung der Annahmen noch eine feste Verteilungsform für die Eintrittswahrscheinlichkeiten einschränkend vorgegeben werden. Denn die Vorgabe einer modellbasierten Methode zur Erzeugung der Fehler um einen mittleren Annahmenverlauf (bzw. um die Punktprognosewerte) eines Modellparameters führt in herkömmlichen probabilistischen Bevölkerungsprognosemodellen zu einer relativ homogenen Annahmenverteilung, deren Annahmenverläufe mehr oder weniger stark um den einen mittleren vorzugebenden Annahmenverlauf (nahezu normalverteilt) schwanken. Neben dem Niveau und dem Verlauf werden dadurch auch die Eintrittswahrscheinlichkeiten der Annahmenverläufe vorgegeben; sofern die zufälligen Abweichungen vom jeweiligen Punktprognosewert den Erwartungswert null und eine festzulegende Varianz erhalten, ist dies gleichbedeutend mit einer Normalverteilung der Eintrittswahrscheinlichkeiten, wobei der mittlere vorzugebende Annahmenverlauf jeweils die höchste Eintrittswahrscheinlichkeit erhält. Wenn ein Prognostiker nun allerdings jedem demographisch interpretierbaren Annahmenverlauf, der jeweils den Vorhersagen einer

reliablen Informationsquelle entspricht, eine Eintrittswahrscheinlichkeit zuordnen möchte, die genau seinen Erwartungen entspricht, so müssen diese subjektiv zugeordnet werden können; eine subjektive Zuordnung der Eintrittswahrscheinlichkeiten gibt demnach keine Verteilungsform für die Eintrittswahrscheinlichkeiten einschränkend vor, sodass z.B. auch zwei heterogenen Annahmenverläufen, die dem unteren und oberen Grenzverlauf einer Annahmenverteilung entsprechen, dieselbe, aber insgesamt höchste Eintrittswahrscheinlichkeit zugeordnet werden kann. Daneben gibt es auch unterschiedliche Auffassungen zur Korrelation der Annahmenwerte eines Modellparameters über das Alter oder/und die Zeit oder/und zwischen den Annahmenwerten verschiedener Modellparameter. Derartige Korrelationsstrukturen sollten im PPPM flexibel, je nach Überzeugung eines Prognostikers, umgesetzt werden können. Außerdem sollten mit dem PPPM vielfältige Prognoseergebnisse auf der Makroebene erzeugt werden können, die auch für diverse bevölkerungsdynamische Analysen eingesetzt werden können. Um beispielsweise die bevölkerungsdynamischen Effekte von Zu- und Fortzügen, die sich in der Alters- und Geschlechtsstruktur wie auch im demographischen Verhalten (untereinander und von den Einheimischen unterscheiden) angemessen berücksichtigen (und analysieren) zu können, sollte im PPPM eine erweiterte Kohorten-Komponenten Methode zur Bevölkerungsprognose verwendet werden, sodass diverse Subpopulationen (wie die Einheimischen, die Zu- und Fortzüge wie auch deren Nachkommengenerationen) mit separaten Annahmen zur Mortalität und Fertilität einzeln prognostiziert werden können. (Damit könnten beispielsweise die Auswirkungen der Zu- und Fortzüge auf die zukünftige Entwicklung der Gesamtbevölkerung analysiert werden.)

Die Modellspezifikation und Implementierung des PPPM werden umfassend in den Kapiteln 9 bis 11 erläutert. Darüber hinaus werden die Besonderheiten des PPPM gegenüber herkömmlichen (probabilistischen) Bevölkerungsprognose(modelle)n anhand einer komplexen Beispielprojektion ausführlich im Kapitel 12 aufgezeigt.

## 5. Generierung von Mortalitätsannahmen

Mortalitätsprognosen sind ein wesentlicher Bestandteil von Bevölkerungsprognosen, obgleich ihre Unsicherheit gegenüber der Prognose der Fertilität und Migration etwas geringer ist; zumal die vergangene Entwicklung der Fertilität und Migration weitaus größeren Schwankungen unterworfen ist als die der Mortalität (vgl. Kapitel 6 und 7). Die Unsicherheit bei der Prognose der Mortalität, Fertilität und Migration hängt dabei u. a. von der Stärke des Einflusses individueller Entscheidungsprozesse auf das jeweilige demographische Verhalten ab, wobei die individuellen Entscheidungen oftmals von einem Prognostiker aufgrund ihrer Komplexität, Privatheit und Instabilität nur schwer eingeschätzt werden können.

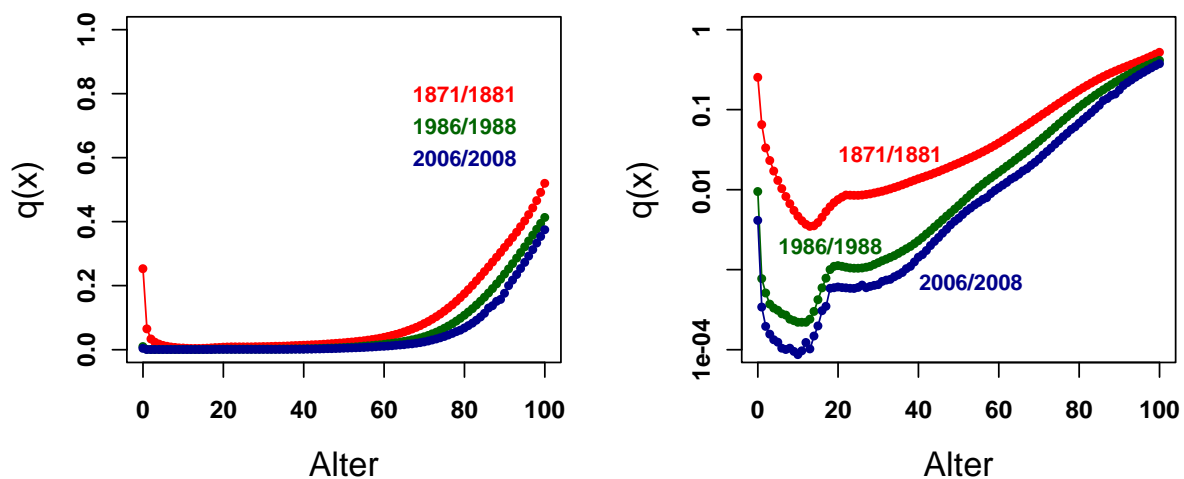


Abbildung 5.1.: Altersspezifische Sterbewahrscheinlichkeiten der Männer in Deutschland aus den Sterbetafeln 1871/81 (rot), 1986/88 (grün) und 2006/08 (dunkelblau) mit linearer (links) und logarithmischer Skala (rechts); Datenquelle: Statistisches Bundesamt

Anhand der Abbildung 5.1 ist zu erkennen, dass der Rückgang der Mortalität zwar relativ stabil ist, jedoch mit dem Alter und der Zeit variiert. Dominierte in der ersten Hälfte des 20. Jahrhunderts der Rückgang der Sterblichkeit in den jungen Altern, setzte ab der zweiten Hälfte des 20. Jahrhunderts verstärkt der Rückgang der Sterblichkeit in den höheren



Altern ein. In diesem Altersbereich liegt derzeit auch das größte Potential für eine weitere Mortalitätsreduktion.

## 5.1. Bestimmung eines geeigneten Parameters zur Mortalitätsprognose

Um die zukünftige Entwicklung der Mortalität exakt zu prognostizieren, muss ein Mortalitätsparameter zur Prognose verwendet werden, der sensibel gegenüber Veränderungen der Mortalität in einzelnen Altern reagiert. Mortalitätsparameter wie die altersspezifische Sterberate oder die altersspezifische Sterbewahrscheinlichkeit erfüllen diese Anforderung, wohingegen der Parameter Lebenserwartung bei Geburt ungeeignet ist, da verschiedene Verteilungen der altersspezifischen Sterbewahrscheinlichkeiten denselben Wert für die Lebenserwartung bei Geburt erzeugen können.<sup>1</sup>

Wenn die Sterblichkeit mit dem Parameter altersspezifische Sterbewahrscheinlichkeit,  $q_x$ , prognostiziert werden soll, darf ihr limitierter Wertebereich, der zwischen 0 und 1 liegt, weder unter- noch überschritten werden. Diese Forderung ist bei einer Extrapolation (vgl. Abschnitt 5.2.1 ab Seite 83) des in einer Basisperiode beobachtbaren Trends über einen längeren Projektionshorizont nicht immer einzuhalten, sodass die untere Grenze bei einem Sterblichkeitsrückgang fälschlicherweise unterschritten werden kann. Um dieses Problem zu beheben, können Transformationen der originalen altersspezifischen Sterbewahrscheinlichkeiten wie z. B. der (natürliche) Logarithmus<sup>2</sup> oder auch der Logit<sup>3</sup> ersatzweise prognostiziert werden. Der Wertebereich für die zu prognostizierenden logarithmierten Sterbewahrscheinlichkeiten liegt zwischen 0 und  $-\infty$  und der für die logit-transformierten Sterbewahrscheinlichkeiten liegt zwischen  $+\infty$  und  $-\infty$ , sodass das Unterschreiten einer festen unteren Grenze bei einer Mortalitätsreduktion ausgeschlossen werden kann. Denn bei der Umrechnung der prognostizierten transformierten in die originalen Sterbewahrscheinlichkeiten fallen wieder alle Werte in den Bereich zwischen 0 und 1 (vgl. Abbildung 5.2).

---

<sup>1</sup>Dennoch wird der Parameter Lebenserwartung bei Geburt häufig in amtlichen Bevölkerungsprognosen als Mortalitätsindikator verwendet.

<sup>2</sup> $\ln q_x$

<sup>3</sup>Der Logit einer altersspezifischen Sterbewahrscheinlichkeit tritt in der Literatur [19, 221] in zwei Varianten auf:  $\text{logit}(q_x) = \ln \frac{q_x}{1-q_x}$  und  $\text{logit}(q_x) = 0,5 \ln \frac{q_x}{1-q_x}$ .

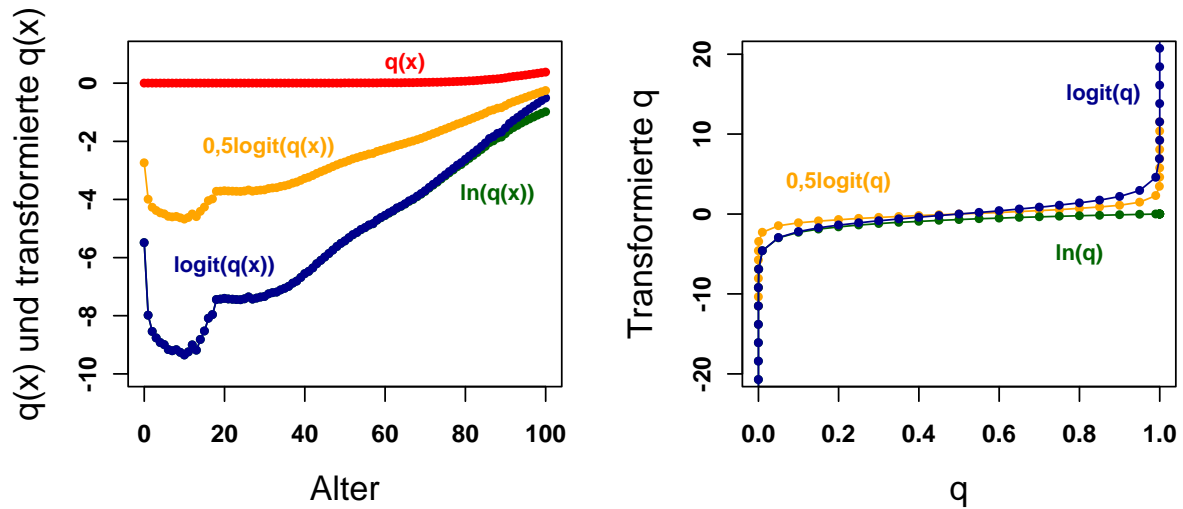


Abbildung 5.2.: Transformation (altersspezifischer) Sterbewahrscheinlichkeiten der Männer in Deutschland aus der Sterbetafel 2006/08: Natürlich logarithmiert (grün), logit-transformiert (dunkelblau), 0,5-logit-transformiert (orange) und original (rot).

## 5.2. Überblick über verschiedene Ansätze zur Mortalitätsprognose

Es gibt eine Vielzahl von Ansätzen zur Prognose der Mortalität, die sich hinsichtlich ihres Prognosegegenstands, ihrer Prognosemethode wie auch hinsichtlich ihrer genutzten Informationsquellen unterscheiden. Demzufolge ist die Wahl des Ansatzes zur Mortalitätsprognose von der Art, dem Umfang und der Qualität der verfügbaren Daten, von der Fragestellung wie auch von der Länge des Projektionshorizonts abhängig [220]. Sind in der Vergangenheit vornehmlich von Demographen und Aktuaren fast ausschließlich trendorientierte Prognosemethoden wie z. B. die Extrapolation historischer Trends zur Mortalitätsprognose herangezogen worden, werden in letzter Zeit neben den demographisch-statistischen Faktoren sukzessive biomedizinische und epidemiologische Faktoren in multikausalen Prognosemethoden berücksichtigt [275, 276].

In der Fachliteratur gibt es keine einheitliche Systematisierung der bestehenden Ansätze zur Mortalitätsprognose. Gleichwohl können — im Hinblick auf die (korrigierende) Berücksichtigung von Expertenmeinung — extrapolative Ansätze von interpolativen, multikausalen und sonstigen Ansätzen zur Mortalitätsprognose unterschieden werden

[11, 30, 33, 97, 128, 220, 221, 275, 276].

In diesem Kapitel sollen nun dieser entwickelten Systematik folgend einige bekannte Ansätze zur Mortalitätsprognose vorgestellt werden, um einen Überblick über die verschiedenen bestehenden Methoden zu geben. Allerdings ist der Bereich der Mortalitätsprognosen derart umfangreich, dass dieser Überblick keinesfalls als erschöpfend bzw. allumfassend, sondern lediglich als Einblick verstanden werden kann. Methoden, die in der Gegenwart besonders häufig verwendet werden, wie beispielsweise der Ansatz von Lee und Carter [50, 157], werden etwas detaillierter vorgestellt.

### 5.2.1. Extrapolative Ansätze

Die Extrapolation ist eine relativ einfache und weit verbreitete Methode zur Mortalitätsprognose. Sie schreibt die in der historischen Zeitreihe des zu prognostizierenden Mortalitätsparameters beobachteten Trends in die Zukunft fort. Damit geht die Extrapolation von der Annahme aus, dass allein die vergangene quantitative Entwicklung eines Mortalitätsparameters dessen zukünftige Entwicklung bedingt und erklärt.

Um die Mortalität extrapolieren zu können, müssen zunächst ein Mortalitätsparameter, eine Extrapolationsmethode wie auch die zu berücksichtigenden Basisperiodenwerte bestimmt werden. Die Wahl eines geeigneten Mortalitätsparameters (wie z. B. die Sterbewahrscheinlichkeit) erfordert einerseits die Entscheidung, auf welchem Aggregationsniveau der Parameter prognostiziert werden soll, und, andererseits, die Entscheidung, ob der Parameter mit originalen oder transformierten (vgl. Abschnitt 5.1 auf Seite 81) Werten prognostiziert werden soll. Darüber hinaus wird mit der Wahl einer Extrapolationsmethode — einfach oder komplex<sup>4</sup> — bestimmt, wie viele vergangene Zeitpunkte berücksichtigt, ob nichtlineare, eng an die historische Entwicklung angepasste Verläufe modelliert und ob stochastische Elemente zum Ausdruck der Unsicherheit eingeschlossen werden. Daneben werden mit der Wahl der Basisperiode die Werte bestimmt, die zur Extrapolation genutzt werden. Folglich ist die in der Basisperiode enthaltene Mortalitätsentwicklung maßgebend für den zukünftigen Mortalitätsverlauf.

---

<sup>4</sup>Eine Übersicht zu Trendextrapolationsmethoden ist z. B. bei Preston et al. [256] zu finden.

## Mathematische und graphische Extrapolation

Pollard [220] unterscheidet grundsätzlich zwischen der mathematischen und der graphischen Extrapolation.

Bei der mathematischen Extrapolation wird ein Trendextrapolationsmodell (wie z. B. ein exponentielles oder ein lineares Modell) an die Basisdaten des zu prognostizierenden Mortalitätsparameters angepasst. Dabei wird der zu prognostizierende Mortalitätsparameter in Abhängigkeit von der Zeit mit einem funktionalen Modell beschrieben. Das Einsetzen zukünftiger Werte für die Zeit in das jeweilige Modell erbringt dann die extrapolierten, zukünftigen Werte des zu prognostizierenden Mortalitätsparameters.

Bei der graphischen Extrapolation wird die meist geglättete, historische Zeitreihe des zu prognostizierenden Mortalitätsparameters manuell *fortgezeichnet*. Anschließend wird der *fortgezeichnete* Verlauf in numerische Daten transformiert, so dass dann die graphisch extrapolierten Daten für den zu prognostizierenden Mortalitätsparameter vorliegen.

## Ansatz von Lee-Carter

Lee und Carter [50, 151, 157] haben 1992 einen Ansatz zur Mortalitätsprognose vorgelegt, der seither als das Standardmodell vielfach genutzt wird. Grundlegend werden mit diesem Ansatz die Sterblichkeitsentwicklungen einer Basisperiode mit einem einfachen Zeitreihenmodell in die Zukunft fortgeschrieben, wobei die Modellparameter mit der Singulärwertzerlegung (oder auch mit einem anderen Regressionsverfahren) geschätzt werden. Das Lee-Carter Modell wird häufig als Vergleichsmodell zum Testen der Prognosegüte (vgl. Kapitel 8) anderer Mortalitätsprognosemethoden herangezogen.

Lee und Carter [157] prognostizieren die logarithmierten, alters- und zeitspezifischen Sterberaten  $m_{x,t}$  mit dem folgenden, an die Werte der Basisperiode<sup>5</sup> anzupassenden Modell (vgl. Formel 5.1).

$$\ln m_{x,t} = \alpha_x + \beta_x k_t + \epsilon_{x,t} \quad (5.1)$$

Der Modellparameter  $\alpha_x$  entspricht der generellen Altersverteilung der Sterberaten. Zur

---

<sup>5</sup>Die Basisperiode enthält die Sterberaten für  $x$  Alter über  $t$  vergangene Jahre, wobei Lee und Carter [157] die Verwendung der längstmöglichen Basisperiode empfehlen.

Bestimmung der Werte von  $\alpha_x$  wird für jedes Alter  $x$  der Durchschnitt der logarithmierten Sterberaten des entsprechenden Alters über den Basiszeitraum gebildet. Demnach variieren die Werte für  $\alpha_x$  mit dem Alter  $x$ , jedoch nicht mit der Zeit  $t$ .

Der Modellparameter  $\beta_x$  misst die Richtung und das Ausmaß der Veränderung der Sterberaten in jedem Alter  $x$ . Die Werte für  $\beta_x$  sind quasi ein Maß für die Standardabweichung der logarithmierten Sterberaten der einzelnen Alter über die Zeit [121], sodass sie angeben, welche Sterberate in welchem Alter sich wie schnell oder langsam im Zusammenspiel mit dem Zeitindex  $k_t$  verändert. Die Werte für  $\beta_x$  können ein positives oder ein negatives Vorzeichen haben, was für einen Rückgang oder Anstieg der Sterblichkeit im jeweiligen Alter steht. Zudem sind auch die Werte für  $\beta_x$  zeitinvariant.

Der Modellparameter  $k_t$  ist der Index für das Mortalitätslevel, der den Rückgang der Sterblichkeit über die Zeit ausdrückt. Die Werte für  $k_t$  entsprechen quasi der Summe der zentrierten logarithmierten Sterberaten über alle Alter in einem Jahr  $t$ .<sup>6</sup> Das Mortalitätslevel  $k_t$  variiert mit der Zeit und ist für alle Alter konstant.

Der Modellparameter  $\epsilon_{x,t}$  ist der (normalverteilte) Fehlerterm des Modells mit einem Mittelwert von 0 und einer konstanten Varianz, der die vom Modell nicht eingefangenen Abweichungen der modellierten von den tatsächlich beobachteten Werten wiedergibt.

Die Werte für  $\beta_x$  und  $k_t$  müssen mit einem geeigneten Schätzverfahren an die Werte der Basisperiode angepasst werden. Lee und Carter [157] schlagen die Methode der Singulärwertzerlegung vor, bei der eine Matrix mit den zentrierten logarithmierten alters- und zeitspezifischen Sterberaten der Basisperiode in ein Produkt aus drei Matrizen mit einer bestimmten Form so zerlegt wird, dass die quadrierten Abweichungen minimal sind<sup>7</sup> (vgl. Formel 5.2).

$$\ln m_{x,t} - \alpha_x = UDV' \quad (5.2)$$

$D$  ist eine diagonale Matrix der Dimension  $x$  mal  $t$ , die die Singulärwerte in absteigender Reihenfolge in der Hauptdiagonalen enthält.  $U$  und  $V'$  sind quadratische Matrizen der

---

<sup>6</sup>Die zentrierte logarithmierte Sterberate ist die Differenz aus der logarithmierten Sterberate in einem Alter  $x$  und dem entsprechenden Wert für  $\alpha_x$ .

<sup>7</sup>Zwei Bedingungen müssen bei diesem iterativen Lösungsverfahren erfüllt werden: Die Summe der  $\beta_x$  muss 1 und die Summe der  $k_t$  muss 0 ergeben.

$$m_{x,t} - \alpha_x = \begin{pmatrix} m_{1,1} - \alpha_1 & m_{1,2} - \alpha_1 & \dots & m_{1,t} - \alpha_1 \\ m_{2,1} - \alpha_2 & m_{2,2} - \alpha_2 & \dots & m_{2,t} - \alpha_2 \\ \vdots & \vdots & \dots & \vdots \\ m_{100+,1} - \alpha_{100+} & m_{100+,2} - \alpha_{100+} & \dots & m_{100+,t} - \alpha_{100+} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_{100+} \end{pmatrix}$$

$$k_1 \qquad k_2 \qquad \dots \qquad k_t$$

Abbildung 5.3.: Bestimmung der zentrierten Sterberaten,  $\ln m_{x,t} - \alpha_x$ , über die Randverteilungen  $k_t$  und  $\beta_x$  im Lee-Carter Modell

Dimension  $x$  mal  $x$  und  $t$  mal  $t$ . Die Werte für  $\beta_x$  werden mit den Werten der ersten Spalte der Matrix  $U$  gleichgesetzt, wohingegen die Werte für  $k_t$  aus dem Produkt der ersten Spalte von  $V$ <sup>8</sup> und dem höchsten Singulärwert von  $D$  bestimmt werden [97, 212]. Die Singulärwertzerlegung kann mit diversen Softwarepaketen wie z. B. R [223] ausgeführt werden. Sie ist ein relativ grobes Verfahren mit dem lediglich Haupttrends der Mortalitätsentwicklung erfasst werden können, sodass subtile, einzelne Aspekte dieser Entwicklung oftmals unberücksichtigt bleiben [121]. Neben der Singulärwertzerlegung können auch andere Verfahren zur Schätzung der Werte für  $\beta_x$  und  $k_t$  wie z. B. die Methode der (Gewichteten) Kleinste-Quadrate-Schätzung herangezogen werden [131, 315].

Nach Girosi und King [97] können die zentrierten Sterberaten ( $\ln m_{x,t} - \alpha_x$ ) aus dem Produkt der jeweiligen Randverteilungswerte, die (in etwa) den Modellparametern  $k_t$  und  $\beta_x$  entsprechen, approximiert werden (vgl. Abbildung 5.3 und Formel 5.3). Demnach kann z. B. die zentrierte Sterberate im Alter 2 im Jahr 1 aus dem Produkt von  $k_1$  und  $\beta_2$  berechnet werden.

$$\ln m_{x,t} - \alpha_x = \beta_x k_t \tag{5.3}$$

Um eine noch bessere Anpassung an die Daten aus der Basisperiode zu erzielen, können die Werte für  $k_t$  zusätzlich an absolute Sterbefallzahlen angepasst werden. Der Aspekt der erneuten Anpassung von  $k_t$  wird jedoch in der Literatur kritisch behandelt [31].

Nachdem alle Modellparameterwerte bestimmt worden sind, wird die für das Mortalitäts-

---

<sup>8</sup>(oder den Durchschnittswerten der zentrierten Sterberaten pro Jahr  $t$ )

level  $k_t$  ermittelte historische Zeitreihe mit einem geeigneten Zeitreihenmodell wie z. B. einem ARIMA-Modell<sup>9</sup> zuerst geschätzt und dann über  $s$  Jahre (bzw. Projektionsintervalle) fortgeschrieben.<sup>10</sup> Diese prognostizierten Werte für das Mortalitätslevel  $k_t$  werden dann in das originale Lee-Carter Modell (vgl. Formel 5.1) eingesetzt, sodass zukünftige Werte für die logarithmierten alters- und zeitspezifischen Sterberaten bestimmt werden können (vgl. Formel 5.4).

$$\ln m_{x,t+s} = \alpha_x + \beta_x k_{t+s} \quad (5.4)$$

Um die Unsicherheit einer solchen Mortalitätsprognose zu quantifizieren, können zudem noch Konfidenzintervalle bestimmt werden.

Die Vorteile des Lee-Carter Modells liegen im Vergleich zu anderen Ansätzen zur Mortalitätsprognose in seinem relativ einfachen Aufbau wie auch in seiner Robustheit. So nutzt das Lee-Carter Modell verhältnismäßig wenig Modellparameter, die weitestgehend alle demographisch interpretiert werden können, sodass der Aufwand der Schätzung der entsprechenden Modellparameterwerte relativ gering ist. Außerdem können mit dem Lee-Carter Modell wesentliche Trends in den Daten der Basisperiode ermittelt und eingefangen werden. Die Prognoseunsicherheit kann zudem mit entsprechenden Konfidenzintervallen quantifiziert werden.

Trotz dieser zahlreichen Vorzüge weist das Lee-Carter Modell aber auch einige Schwächen auf. So können durch die Verwendung roher, ungeglätteter Sterberaten Verzerrungen bei der Schätzung der Modellparameter entstehen, wenn Ausreißer oder auch zufällige erratische Sprünge in den Originaldaten als vermeintliche Trends fehlinterpretiert und als solche in das Modell eingebaut werden. Darüber hinaus variiert der Rückgang der Sterblichkeit in der Realität mit dem Alter *und* der Zeit (vgl. Abbildung 5.1). Allerdings kann mit dem Lee-Carter Modell nur ein sich zeitlich verändernder, aber für alle Alter konstanter Sterblichkeitsrückgang über die  $k_t$  modelliert werden. Das Ausmaß der relativen Veränderung der Sterblichkeit zwischen den Altern bleibt dabei über den gesamten Projektionshorizont durch die  $\beta_x$  unverändert. Eine wirklichkeitsgetreue Interaktion zwischen dem Alter und der Zeit würde erst erreicht werden, wenn sich die  $k_t$  auch über das Alter und die  $\beta_x$  auch mit der Zeit verändern könnten. Diese zusätzliche Variabilität würde die Prognose eines über den Projektionshorizont sinkenden Sterblichkeitsrückgangs in einem Alter (wie in der

---

<sup>9</sup>AutoRegressiveIntegratedMovingAverage nach Box und Jenkins [36]

<sup>10</sup>Lee und Carter [157] schlagen dafür einen Random Walk mit Drift vor.

Vergangenheit bei den jungen Altern) und eines gleichzeitig ansteigenden Sterblichkeitsrückgangs in einem anderen Alter (wie in der Vergangenheit bei den hohen Altern) ermöglichen. Die stark vereinfachende Annahme der Unabhängigkeit der Alters- und Zeitachse wirkt sich auch auf die Eignung des Lee-Carter Modells zur Langzeitprognose der Mortalität aus: Je länger der Projektionshorizont ist, desto unplausibler ist diese Annahme. Demnach ist das Lee-Carter Modell eher für Kurzzeitprognosen geeignet, zumal es auch ausschließlich die vergangene empirische Entwicklung der Sterblichkeit extrapoliert und keine zusätzlichen Informationen (wie z. B. aus der Medizin) berücksichtigt. Zudem kann das Lee-Carter Modell die Prognoseunsicherheit unterschätzen, da bei der Konstruktion der Konfidenzintervalle nur die Fehler bei der Prognose von  $k_t$  berücksichtigt werden. Unberücksichtigt bleiben z. B. die nicht unerheblichen Fehler bei der Schätzung der (anderen) Modellparameter wie auch bei der Anpassung des Modells an tatsächlich beobachtete Daten [50, 131]. Schließlich erweist sich auch die Nutzung der längsten verfügbaren historischen Zeitreihe zur Schätzung des Lee-Carter Modells als problematisch. Diese Forderung bezieht sich wahrscheinlich auf Box und Jenkins [36], die ein Minimum von 50 Zeitreihenwerten für eine komplexe Zeitreihenanalyse vorgeben. Jedoch hat sich bei vergleichenden Anwendungen des Lee-Carter Modells gezeigt, dass eine kürzere Basisperiode oftmals genauere Ergebnisse hervorbringt, da dann die zuletzt beobachteten Entwicklungen ein größeres Gewicht erhalten [31, 33].

Aufgrund seiner zahlreichen Vorzüge und bekannten und dadurch behebbaren Schwächen ist das Lee-Carter Modell seit seiner Einführung im Jahr 1992 von vielen Wissenschaftlern in der originalen Form, in leicht abgeänderten Varianten oder auch in erweiterten Fassungen zur Mortalitätsprognose verwendet worden [31, 32, 121, 118, 131, 151, 158, 167, 283]. Die aus dem Lee-Carter Modell hervorgegangenen Varianten unterscheiden sich grundsätzlich in der Wahl der Basisperiode, der erneuten Anpassung des Mortalitätslevels  $k_t$  und der Nutzung von geschätzten oder tatsächlich beobachteten Werten für die Sterberaten im aktuell verfügbaren Jahr.<sup>11</sup>

**Die Erweiterung von Hyndman und Ullah** Hyndman und Ullah [118] haben 2007 das Lee-Carter Modell entscheidend erweitert, indem sie etliche seiner Schwächen behoben haben. So verwenden sie ausschließlich geglättete Sterberaten aus der Basisperiode zur Schätzung der Modellparameter. Dadurch können Ausreißer und zufällige erratische Sprünge in den Basisdaten abgemildert werden anstatt sie wie im originalen Lee-Carter Modell als Trend

---

<sup>11</sup>Hierbei handelt es sich um so genannte Jump-off Raten.



fehlzuinterpretieren und ins Modell einzubinden. Außerdem haben Hyndman und Ullah die Modellgleichung dermaßen umgewandelt, dass mit den  $k$  Sets der Basisfunktionen  $\phi_{k,x}$  und den entsprechenden Zeitreihenprozessen  $\beta_{t,k}$  einzelne Altersbereiche separat modelliert und prognostiziert werden können (vgl. Formel 5.5).<sup>12</sup>

$$\ln m_{x,t} = \alpha_x + \sum_{k=1}^K \beta_{t,k} \phi_{k,x} + \epsilon_{x,t} \quad (5.5)$$

Dabei werden die Basisfunktionen  $\phi_{k,x}$ , die eine hohe Korrelation der Sterberaten bestimmter Alter  $x$  anzeigen, mit einer robusten Hauptkomponentenanalyse ermittelt. Der historische Verlauf dieser relativ hoch korrelierenden Altersbereiche wird dann mit dem jeweiligen Zeitreihenprozess  $\beta_{t,k}$  nachgebildet und anschließend mit einem geeigneten Zeitreihenmodell prognostiziert.<sup>13</sup> Eine weitere entscheidende Verbesserung gegenüber dem originalen Lee-Carter Modell betrifft die Quantifizierung der Unsicherheit durch die Konstruktion der Konfidenzintervalle. Berücksichtigen Lee und Carter lediglich die Fehler bei der Prognose des Mortalitätslevels  $k_t$ , berücksichtigen Hyndman und Ullah die Unsicherheit in Form der Varianz aller Modellkomponenten, also auch von  $\alpha_x$ ,  $\beta_x$  (bzw.  $\phi_{k,x}$ ),  $k_t$  (bzw.  $\beta_{t,k}$ ) und von  $\epsilon_{x,t}$ .

Auch de Jong und Tickle [121] haben 2006 das Lee-Carter Modell erweitert. Ihr erweitertes Modell weist ähnliche Merkmale wie das von Hyndman und Ullah auf, allerdings verwenden sie andere Methoden zu deren Umsetzung. So nutzen de Jong und Tickle geglättete Basisdaten, jedoch verwenden sie zur Schätzung der Modellparameter andere Schätzverfahren wie z.B. die Kleinste-Quadrate Methode oder die Maximum-Likelihood Methode. Zudem führen sie auch einen mit der Zeit variierenden Mortalitätstrend ein, der für verschiedene Altersbereiche separat modelliert werden kann.

Die beschriebenen Erweiterungen von Hyndman und Ullah und von de Jong und Tickle sollen eine bessere Anpassung gerade auch an instabile Mortalitätsentwicklungen erzielen. Insbesondere bei der Erweiterung von Hyndman und Ullah können durch die flexibel anzupassenden  $k$  Sets der Basisfunktion und der jeweiligen Zeitreihenprozesse besondere

---

<sup>12</sup>Der Modellparameter  $\beta_{t,k}$  entspricht in etwa dem Mortalitätslevel  $k_t$  aus dem originalen Lee-Carter Modell.

<sup>13</sup>Hyndman und Ullah [118] ermitteln bei einer Mortalitätsprognose für Frankreich mit der Basisperiode von 1899 bis 2001 insgesamt vier Basisfunktionen, die eine relativ hohe Korrelation für die Sterblichkeit bei Kindern ( $\phi_{1,x}$ ), bei jungen Erwachsenen im Verhältnis zu über 60-Jährigen ( $\phi_{2,x}$ ) und bei sehr Alten ( $\phi_{3,x}$ ) abbilden. Die vierte Basisfunktion wird von Hyndman und Ullah als zu komplex eingestuft, um sie entsprechend interpretieren zu können.

Sterblichkeitsentwicklungen in bestimmten Altersbereichen modelliert werden. Generell unterscheiden sich die Varianten und Erweiterungen in der Genauigkeit ihrer Ergebnisse kaum, jedoch sind sie genauer als die Ergebnisse des originalen Lee-Carter Modells [31, 33].

Das Lee-Carter Modell wie auch seine Varianten und Erweiterungen können zudem nicht nur zur Prognose von altersspezifischen Sterberaten, sondern auch zur Prognose von alters- und todesursachenspezifischen Sterberaten verwendet werden (vgl. Abschnitt 5.2.3 ab Seite 102) [52, 97, 314].

### **Mortalitätsgesetze**

Ein Mortalitätsgesetz beschreibt empirische Regelmäßigkeiten in der Verteilung der Sterblichkeit über das Alter mit einer mathematischen Funktion. Die Zahl der dafür zu verwendenden Modellparameter variiert (bisher) zwischen 2 und 9. Bei der Entwicklung eines Mortalitätsgesetzes sind zwei Kriterien zu beachten. Einerseits sollte das Mortalitätsgesetz eine Sterblichkeitsverteilung generieren, die eine hohe Anpassungstreue an die Originaldaten hat. Andererseits sollten die zu verwendenden Modellparameter demographisch sinnvoll interpretierbar sein.

Prinzipiell unterscheiden sich bestehende Mortalitätsgesetze darin, ob sie eine oder mehrere ursächliche Komponenten zur Modellierung der Sterblichkeit enthalten, ob sie die gesamte Altersspanne oder nur einen bestimmten Altersbereich nachbilden und ob sie auf einer Theorie basieren oder einfach die Sterblichkeit in Abhängigkeit von der Zeit über ein entsprechendes Modell modellieren [276].

Bei der Verwendung von Mortalitätsgesetzen zur Mortalitätsprognose werden Daten zur altersspezifischen Sterblichkeit aus der Basisperiode mit dem jeweiligen Mortalitätsgesetz nachgebildet. Dabei entsteht für jeden Modellparameter (des jeweiligen Mortalitätsgesetzes) eine Zeitreihe, die dann fortgeschrieben bzw. mit einer geeigneten Methode extrapoliert werden kann. Diese prognostizierten Modellparameterwerte werden dann in die Funktionsgleichung des Mortalitätsgesetzes eingesetzt, um zukünftige Werte für den zu prognostizierenden Mortalitätsparameter wie z. B. die altersspezifischen Sterbewahrscheinlichkeiten zu erhalten.

Bereits ab der ersten Hälfte des 18. Jahrhunderts ist der Versuch unternommen worden, die Entwicklung der Höhe der Sterblichkeit und ihre Altersverteilung mit einer mathematischen

Funktion zu formulieren. Erste Klassiker sind Mortalitätsgesetze u. a. von de Moivre aus dem Jahr 1725, von Gompertz aus dem Jahr 1825 [98], von Makeham aus dem Jahr 1860, von Thiele aus dem Jahr 1872 [278] oder auch von Opperman aus dem Jahr 1870 [105, 205, 221].

Das Mortalitätsgesetz von Gompertz [98] aus dem Jahr 1825 basiert auf der Annahme, dass die Sterblichkeit mit dem Alter exponentiell steigt (vgl. Formeln 5.6 und 5.7).

$$q_x = Bc^x \quad (5.6)$$

$$\ln q_x = \ln B + x \ln c \quad (5.7)$$

Demzufolge wird die Sterbewahrscheinlichkeit  $q$  im Alter  $x$  von Gompertz mit zwei Konstanten bzw. Faktoren ursächlich erklärt: Einerseits mit der generellen Chance zu sterben ( $B$ ) und andererseits mit der physischen Degeneration ( $c$ ), die mit dem Alter  $x$  zunimmt [285].<sup>14</sup> Das Mortalitätsgesetz von Gompertz gilt ausschließlich für die Sterblichkeit im Erwachsenenalter, da es durch seine inherente Annahme des exponentiellen Anstiegs der Sterblichkeit mit dem Alter weder die Säuglings- und Kindersterblichkeit noch den Unfallbuckel nachbilden kann. Jedoch gibt es auch eine obere Altersgrenze für die Eignung des Mortalitätsgesetzes von Gompertz, da die Sterblichkeit in den ganz hohen Altern nicht mehr exponentiell, sondern nur noch verlangsamt steigt (vgl. Abschnitt 5.2.3 ab Seite 98). Demnach gilt das Mortalitätsgesetz von Gompertz in etwa für den Altersbereich der 25 bis 75-Jährigen, wohingegen er es für die Alter 20 bis 60 ausgewiesen hat [205, 221].

In der Tradition von Thiele [278], der erstmals 1872 vorgeschlagen hat, mit einem Mortalitätsgesetz die Sterblichkeit über alle Alter mit drei additiven Termen, die für die Sterblichkeit der Kinder, der mittleren Erwachsenenalter und der hohen Alter stehen, zu modellieren, haben Heligman und Pollard [105] 1980 und Rogers und Little [239] 1994 zwei populäre Mortalitätsgesetze der näheren Gegenwart konzipiert. Demnach haben Heligman und Pollard ein Mortalitätsgesetz entworfen, das mit drei additiv verbundenen Komponenten und acht verschiedenen Modellparametern die Mortalität der Kindheit, des Erwachsenenalters und die Unfallsterblichkeit im mittleren Alter funktional modelliert. Alle

---

<sup>14</sup>Zur Formulierung dieses Mortalitätsgesetzes hat Gompertz Daten zur Sterblichkeit nach dem Alter von Personen im Erwachsenenalter aus den verschiedenen Ländern und Städten wie Schweden, Carlisle und Northampton genauer auf regelmäßige Strukturen analysiert.

acht Modellparameter können demographisch sinnvoll interpretiert werden. Das multiexponentielle Modell von Rogers und Little basiert hingegen auf fünf additiv miteinander verbundenen Komponenten mit insgesamt 13 verschiedenen Modellparametern. Die Besonderheit dieses multiexponentiellen Modells besteht in dessen allgemeiner Formulierung, die es ermöglicht, Altersverteilungen verschiedener demographischer Parameter aus der Fertilität, Mortalität und Migration flexibel zu modellieren. Je nachdem, welche Modellparameter des multiexponentiellen Modells mit Werten gleich oder ungleich Null belegt werden, können damit u-förmige, ein- und zweigipflige Wahrscheinlichkeitsfunktionen konstruiert werden. Für die Modellierung der Mortalität mit dem multiexponentiellen Modell ist die Verwendung von vier Komponenten mit insgesamt 9 Modellparametern geeignet. Damit enthält das Mortalitätsgesetz von Rogers und Little die höchste Zahl eingeschlossener Modellparameter. Jedoch kann nicht jeder dieser 9 Modellparameter demographisch sinnvoll interpretiert werden. Zudem weist das multiexponentielle Modell im Vergleich zu Modellen, die auf die Modellierung eines demographischen Parameters (wie z.B. das Mortalitätsgesetz von Heligman und Pollard) spezialisiert sind, eine geringere Anpassungstreue an die originalen Daten auf. Neben der Modellierung der altersspezifischen Sterberaten und -wahrscheinlichkeiten können mit dem multiexponentiellen Modell auch todesursachenspezifische Sterberaten und -wahrscheinlichkeiten nach dem Alter modelliert werden [30, 185, 186].

In der Vergangenheit sind sowohl Wachstumsgesetze für die Bevölkerung als auch für die Mortalität mit großer Überzeugung gesucht und unterstellt worden. Jedoch haben bereits die Vertreter der ersten modernen Ansätze zur Bevölkerungsprognose (vgl. Abschnitt 3.1.2 ab Seite 27) Zweifel an lang anhaltenden Mortalitätsgesetzen geäußert [303]. Zwar können Mortalitätsgesetze aufgrund der verfügbaren Rechnerleistung relativ einfach praktisch angewendet werden, allerdings kann es sich als schwierig erweisen, überhaupt ein geeignetes Mortalitätsgesetz zu finden, das die Originaldaten der Basisperiode mit einer hohen Anpassungstreue nachbildet. Deshalb werden Mortalitätsgesetze in der Gegenwart auch eher selten zur Mortalitätsprognose herangezogen. Zumal sie im Vergleich zu anderen Methoden relativ viele Modellparameter einschließen, ohne dabei die Prognosegüte signifikant zu erhöhen. So prognostiziert z.B. das Lee-Carter Modell [50, 157] (vgl. Abschnitt 5.2.1 ab Seite 84) die Sterblichkeit mit nur 3 Modellparametern, die alle sinnvoll interpretiert werden können, wohingegen das Modell von Little und Rogers [239] 9 Modellparameter einschließt, die nicht alle demographisch sinnvoll interpretiert werden können. Folglich sind Mortalitätsgesetze aus der Gegenwart in ihrem Aufbau komplexer, aber in ihrer Prognose nicht signifikant genauer

als vergleichbare Standardmodelle zur Mortalitätsprognose. Zudem nutzen Mortalitätsgesetze ausschließlich die vergangene Sterblichkeitsentwicklung bei der Prognose der Sterblichkeit, sodass auch sie keine zusätzlichen Informationen (wie z. B. aus der Medizin) berücksichtigen.

### Relationales Modell von Brass

Das relationale Modell zur Mortalitätsprognose von Brass aus den 1970ern [37, 38] prognostiziert einen (logit-transformierten)<sup>15</sup> Sterbetafelverlauf  $Y_x$  nach dem Alter  $x$  über einen linearen funktionalen Zusammenhang mit einem (logit-transformierten) Standard- oder Modellsterbetafelverlauf  $Y_x^*$  (vgl. Formeln 5.8 und 5.9 und Abschnitt 5.2.2 ab Seite 95).

$$Y_x = \alpha + \beta Y_x^* \quad (5.8)$$

$$Y_x = \text{logit}(1 - l_x) = 0,5 \ln \frac{1 - l_x}{l_x} \quad (5.9)$$

Zur Bestimmung dieses linearen funktionalen Zusammenhangs werden zwei Modellparameter,  $\alpha$  und  $\beta$ , und ein logit-transformierter Standardsterbetafelverlauf,  $Y_x^*$ , benötigt. Der Standardsterbetafelverlauf gibt einen allgemeinen Sterbetafelverlauf vor, der durch die beiden Modellparameter  $\alpha$  und  $\beta$  im Sterblichkeitsniveau und in der Verteilung der Sterblichkeit über das Alter verändert werden kann [104, 197, 221, 273].

Zur Mortalitätsprognose werden logit-transformierte Sterbetafelverläufe aus einer Basisperiode mit dem relationalen Modell von Brass an einen logit-transformierten Standardsterbetafelverlauf angepasst. Brass [37] schlägt vor, den zum Zeitpunkt der Prognoseerstellung aktuellen Sterbetafelverlauf als Standardsterbetafelverlauf zu verwenden. Bei der Anpassung des relationalen Modells an die Daten der Basisperiode entsteht jeweils eine Zeitreihe für die Modellparameter  $\alpha$  und  $\beta$ , die dann mit einem geeigneten Verfahren extrapoliert werden kann. Anschließend werden diese prognostizierten Werte für die beiden Modellparameter  $\alpha$  und  $\beta$  in das relationale Modell von Brass eingesetzt, um zukünftige Werte der logit-transformierten Sterbetafelverläufe für die einzelnen Projektionsintervalle des Projektionshorizonts zu erhalten.

Das relationale Modell von Brass ist ein relativ einfaches Modell, dass dazu geeignet ist, grobe Veränderungen im Sterblichkeitsverlauf zwischen zwei Bevölkerungen über die Zeit

---

<sup>15</sup>(vgl. Abschnitt 5.1 auf Seite 81)

mit nur zwei Modellparametern zu beschreiben [220]. Doch je größer die Abweichungen des beobachteten Sterbetafelverlaufs vom Standardsterbetafelverlauf sind, desto weniger wird die Annahme ihres linearen Zusammenhangs erfüllt. Demnach kann das relationale Modell von Brass nicht alle in der Realität auftretenden Variationen im Sterblichkeitsniveau und in der Sterblichkeitsverteilung über das Alter einfangen [197].

Das originale relationale Modell von Brass ist einige Male um zusätzliche Modellparameter erweitert worden, um eine noch bessere Anpassung an die Originaldaten der Basisperiode und eine noch höhere Vorhersagekraft bei der Mortalitätsprognose zu erreichen [104, 197].

### Age-Period-Cohort Modelle

Age-Period-Cohort Modelle (kurz APC Modelle) passen meist ein Regressionsmodell mit drei additiv verbundenen Koeffizienten, die Effekte im Alter, in der Periode und in der Kohorte repräsentieren, an logarithmierte alters- und zeitspezifische Sterberaten aus der Basisperiode an, um diese dann damit zu prognostizieren. Jedoch können die Koeffizienten nicht eindeutig identifiziert werden, da es anstatt einer einzigen unendlich viele Lösungen für ein APC Modell gibt [58, 276]. Generell werden APC Modelle eher selten zur Mortalitätsprognose herangezogen, jedoch werden sie im Bereich der Epidemiologie zur Analyse der Sterblichkeit einzelner Todesursachen verwendet [52, 57, 58].

#### 5.2.2. Ansätze mit Zielverteilungen

Neben der Extrapolation gibt es Ansätze zur Mortalitätsprognose, die eine Sterblichkeitsverteilung über das Alter für ein bestimmtes Zieljahr fest vorgeben. Die Auswahl bzw. Generierung dieser Zielverteilung bestimmt den konkreten Ansatz. Das Erreichen der jeweils gewählten Zielverteilung kann mit verschiedenen Methoden wie z. B. der Extrapolation (vgl. Abschnitt 5.2.1 ab Seite 83), Interpolation,<sup>16</sup> Expertenmeinung oder auch mit einer Mischung diverser Methoden erfolgen.

#### Fortgeschrittenere Bevölkerung als Referenz

Einer dieser Ansätze mit Zielverteilungen geht von der Annahme aus, dass die Sterblichkeit der zu betrachtenden Bevölkerung der Sterblichkeit einer fortgeschrittenen Bevölkerung

---

<sup>16</sup>Mit der Methode der Interpolation können Werte zwischen zwei vorgegebenen Punktwerten (wie hier die Werte der Start- und der Zielverteilung) funktional bestimmt werden.

folgt. Dabei kann die Sterblichkeit der zu betrachtenden Bevölkerung dem gleichen Entwicklungspfad folgen (wie ihn die Sterblichkeit der fortgeschritteneren Bevölkerung durchlaufen hat) oder sie kann sich dem Sterblichkeitsniveau einer ausgewählten Zielverteilung auf einem eigenen Entwicklungspfad anpassen.

Für eine konkrete Anwendung dieses Ansatzes muss zunächst eine geeignete fortgeschrittenere Bevölkerung bestimmt werden wie auch der Abstand zwischen der Sterblichkeitsentwicklung der zu betrachtenden und der ausgewählten fortgeschritteneren Bevölkerung. Dafür können sowohl sozioökonomische und kulturelle Faktoren wie auch der Entwicklungsstand der Medizin oder auch die Verteilung der Haupttodesursachen vergleichend in beiden Bevölkerungen betrachtet werden. Anschließend muss der Entwicklungspfad der Sterblichkeit der zu betrachtenden Bevölkerung genau festgelegt werden. Hierfür gibt es keine feste Vorgabe. Der prognostizierte Sterblichkeitsverlauf der zu betrachtenden Bevölkerung kann z. B. zeitversetzt exakt dem Sterblichkeitsverlauf der fortgeschritteneren Bevölkerung entsprechen oder mit einer Interpolation zwischen Start- und Zielverteilung der Sterblichkeit ermittelt werden [220, 256].

Die praktische Umsetzung dieses Ansatzes zur Mortalitätsprognose birgt einige Probleme. So hat der Prognostiker kaum feste Vorgaben an die er sich bei einer Prognose mit diesem Ansatz halten kann. Dabei ist die Wahl einer geeigneten fortgeschritteneren Bevölkerung schwierig wie auch die Prognose der Entwicklung des Abstands der Sterblichkeitsverteilungen beider Bevölkerungen. Diese Problematik wird noch verstärkt durch die unterschiedlichen Gründe, die meistens für die Sterblichkeitsentwicklung beider Bevölkerungen verantwortlich waren, sind und sein werden. Positiv hervorzuheben ist die Möglichkeit bei diesem Ansatz auch zusätzliche Informationen (wie z. B. aus der Medizin) indirekt in die Prognose einfließen zu lassen.

### **Modellsterbetafel als Referenz**

Ein weiterer Ansatz zur Mortalitätsprognose dieser Kategorie nutzt eine Modellsterbetafel als Zielverteilung für die zu betrachtende Bevölkerung.

Ein System von Modellsterbetafeln, das derzeit am häufigsten verwendet wird, ist von Coale und Demeny [59, 60] in der zweiten Hälfte des 20. Jahrhunderts entwickelt worden. Das Modellsterbetafelsystem von 1983 unterscheidet jeweils 25 verschiedene Sterblichkeitsniveaus

in insgesamt vier unterschiedlichen Sterblichkeitsverteilungen oder Mustern (Nord, Süd, Ost und West). Jedes dieser 25 Sterblichkeitsniveaus pro Muster wird durch den Mortalitätsparameter fernere Lebenserwartung im Alter 10,  $e_{10}$ , repräsentiert.<sup>17</sup> Deutschland fällt in das Muster Ost, das durch eine relativ hohe Säuglings- und Kleinkindsterblichkeit und eine relativ hoch ansteigende Sterblichkeit ab dem Alter 50 gekennzeichnet ist.

Für eine Mortalitätsprognose mit diesem Ansatz muss der Prognostiker sowohl das zukünftige Niveau des Mortalitätsparameters  $e_{10}$  als auch das Muster der Sterblichkeit der zu betrachtenden Bevölkerung bestimmen. Es gibt keine festen Vorgaben für diese beiden Arbeitsschritte. Demnach kann  $e_{10}$  für das Zieljahr mit jeder geeigneten Methode wie z. B. der Extrapolation, der Interpolation oder auch der Expertenmeinung prognostiziert werden. Bei der Wahl des Musters der Sterblichkeit kann sich der Prognostiker an der vergangenen Entwicklung der Sterblichkeitsverteilung oder auch an der eines Landes mit ähnlichen mortalitätsbedingenden Strukturen orientieren. Wenn der zukünftige Wert für  $e_{10}$  und das zukünftige Muster der Sterblichkeit der zu betrachtenden Bevölkerung bestimmt sind, kann die entsprechende Modellsterbetafel mit altersspezifischen Sterbewahrscheinlichkeiten als Zielverteilung der Sterblichkeit eingesetzt werden [220, 221, 273].

Die Vereinten Nationen [284] haben ein anderes Modellsterbetafelsystem vorgeschlagen, das gerade auch Sterblichkeitsentwicklungen für Entwicklungsländer widerspiegeln soll. Sie unterscheiden die vier Regionen Lateinamerika, Chile, Südasien und Ferner Osten sowie ein generelles Muster. Jedes Sterblichkeitsniveau einer Region wird durch den Mortalitätsparameter Lebenserwartung bei Geburt,  $e_0$ , repräsentiert. Die Vorgehensweise für eine Mortalitätsprognose bleibt auch mit diesen Modellsterbetafeln unverändert [221, 273].

Problematisch ist die Anwendung dieses Ansatzes, wenn das zukünftige Sterblichkeitsniveau der zu betrachtenden Bevölkerung nicht im zu verwendenden Modellsterbetafelsystem enthalten ist, sprich wenn das zukünftige Sterblichkeitsniveau der zu betrachtenden Bevölkerung die enthaltenen Sterblichkeitsniveaus des Modellssterbetafelsystems unter- oder überschreitet. Ähnlich verhält es sich mit der zukünftigen Sterblichkeitsverteilung der zu betrachtenden Bevölkerung; ist diese im Modellsterbetafelsystem nicht enthalten, so wird die Prognose ungenau [197, 221]. Um diese Probleme zu umgehen wird dieser Ansatz meist

---

<sup>17</sup>Die Sterblichkeitsniveaus eines Musters reichen von einer Lebenserwartung bei Geburt von 20 bis 80 in 2,5er Schritten, so dass daraus insgesamt 25 Niveaus resultieren.



zur Prognose der Sterblichkeit für Bevölkerungen von Entwicklungsländern verwendet. Eine weitere Schwäche liegt in der Verwendung des Mortalitätsparameters Lebenserwartung bei Geburt ( $e_0$ ) oder im Alter 10 ( $e_{10}$ ) als Repräsentant für eine Sterblichkeitsverteilung und ein Sterblichkeitsniveau (vgl. Abschnitt 5.1 auf Seite 81).

### Optimale Sterbetafel als Referenz

Anstatt einer Modellsterbetafel nutzt ein weiterer Ansatz zur Mortalitätsprognose eine so genannte optimale Sterbetafel als Zielverteilung der Sterblichkeit einer zu betrachtenden Bevölkerung. Damit verbunden ist die Frage, wie weit die Sterblichkeit in jedem Alter tatsächlich sinken kann (vgl. Abschnitt 5.2.3 ab Seite 98).

Bei diesem Ansatz wird demnach eine optimale Sterbetafel erstellt, indem die höchsten, überhaupt möglichen Verbesserungen in der Sterblichkeit angenommen werden. So kann eine optimale Sterbetafel z.B. aus den niedrigsten, in den verschiedenen Ländern der Erde vorkommenden altersspezifischen Sterbewahrscheinlichkeiten oder aber auch aus den altersspezifischen Sterbewahrscheinlichkeiten des Landes mit der geringsten Sterblichkeit bestehen.<sup>18</sup> Nachdem diese optimale Zielverteilung der Sterblichkeit für die zu betrachtende Bevölkerung festgelegt worden ist, muss anschließend der Entwicklungspfad der Sterblichkeit von der Start- zur Zielverteilung mit einer geeigneten Methode wie z.B. der Interpolation bestimmt werden [220].

Ähnlich wie bei den anderen beiden Ansätzen mit einer Zielverteilung für die Sterblichkeit gibt es kaum feste Vorgaben zur Durchführung der einzelnen Arbeitsschritte für einen Prognostiker. Zudem ist hier die Erstellung der Zielverteilung in Form einer optimalen Sterbetafel in der Zukunft besonders vage, da entweder ausschließlich theoretische Überlegungen oder auch die niedrigsten Sterberaten in einzelnen Altern aus verschiedenen Bevölkerungen zum Zeitpunkt der Prognoseerstellung herangezogen werden können. Positiv ist jedoch die mögliche Berücksichtigung zusätzlicher Informationen wie beispielsweise erklärender Theorien zum Eingriff in den menschlichen Alterungsprozess in der Zukunft herauszustellen.

---

<sup>18</sup>Daneben können z.B. auch die beobachteten Sterbewahrscheinlichkeiten *optimaler* Subpopulationen zur Konstruktion einer optimalen Sterbetafel verwendet werden, die besonders günstigen, mortalitätssenkenden und damit lebensverlängernden Lebensbedingungen ausgesetzt sind. So ist beispielsweise die Sterblichkeit aktiver kalifornischer Mormonen deutlich geringer als die der (weißen) Allgemeinbevölkerung in den USA [80, 183].

### 5.2.3. Multikausale Ansätze

#### Mortalitätsprognose mit Überlegungen zur Langlebigkeit

Es gibt seit langer Zeit eine wissenschaftliche Diskussion über die Langlebigkeit von Menschen in der Demographie. Dabei spielt die Frage nach der maximal zu erreichenden Lebensdauer (für ein Individuum oder im Durchschnitt für eine ganze Bevölkerung) eine zentrale Rolle. Mindestens zwei Positionen sind in dieser Diskussion zu unterscheiden: So schätzt eine Gruppe von Wissenschaftlern, die sich bei Mortalitätsprognosen vornehmlich auf evolutionsbiologische Argumente stützt, ein Ansteigen der Lebenserwartung bei Geburt auf über 85 bei Ausbleiben sensationeller medizinischer Innovationen zum Eingreifen in den menschlichen Alterungsprozess als unwahrscheinlich ein. Darüber hinaus schließen sie sogar einen Rückgang der zukünftigen Lebenserwartung bei Geburt aufgrund bestimmter aktueller Gesundheitstrends (wie z. B. Adipositas) in den Bevölkerungen der Industrieländer nicht ganz aus [49, 206]. Im Gegensatz dazu hält die Gruppe von Wissenschaftlern, die sich vornehmlich auf die empirisch messbare Sterblichkeitsentwicklung bezieht, die Fortführung des in der Vergangenheit beobachteten, ansteigenden Trends der Lebenserwartung bei Geburt in der Zukunft für wahrscheinlich [312]. Damit kann die Lebenserwartung bei Geburt noch in diesem Jahrhundert auf über 100 ansteigen [201]. Die Vertreter dieser Positionen stimmen darin überein, dass die Sterblichkeit in den Altern in der Zukunft sinken kann, jedoch unterscheiden sie sich in ihren Annahmen über das Ausmaß, den Verlauf, die Geschwindigkeit und die dahinterliegenden Prozesse dieser Mortalitätsreduktion.<sup>19</sup>

Die Empiriker gehen bei einer Mortalitätsprognose davon aus, dass die in der Vergangenheit observierten Trends in die Zukunft übertragen und extrapoliert werden sollten [201, 296, 297, 301]. So ist die Lebenserwartung bei Geburt in den Ländern mit der niedrigsten Mortalität in den letzten 160 Jahren linear um 3 Monate pro Jahr oder auch um 2,5 Jahre pro Dekade gestiegen. Bei Fortschreibung dieses linearen Trends würde eine Lebenserwartung bei Geburt von 100 bereits nach 60 Jahren erreicht werden. Der unverminderte Anstieg der Lebenserwartung bei Geburt in der Zukunft sei vornehmlich durch die seit 1950 bzw. verstärkt seit 1970 einsetzende Reduktion der Sterblichkeit in den höchsten Altern, also ab

---

<sup>19</sup>Nachfolgend werden die Vertreter dieser beiden Positionen als *Wissenschaftler, die sich vornehmlich auf evolutionsbiologische Argumente stützen*, bzw. als *Empiriker* bezeichnet; diese Bezeichnung stellt keine klare Trennung nach ausschließlich zu verwendenden Faktoren zur Erklärung der Sterblichkeitsentwicklung für die Vertreter der jeweiligen Position dar; vielmehr dient sie in dieser Arbeit zur einfachen Unterscheidung der Vertreter beider Positionen.

dem Alter 80, durch z. B. biomedizinische Fortschritte zu erklären. Die Empiriker sprechen sich sowohl gegen Unsterblichkeit als auch gegen eine (naheliegende) obere Grenze für die Lebenserwartung bei Geburt aus. Ihre Annahme begründen sie neben dem beobachteten linearen Langzeittrend damit, dass in der Vergangenheit schon oft ultimative Level für die Lebenserwartung bei Geburt aufgestellt und kurze Zeit nach ihrer Veröffentlichung bereits erreicht und auch überschritten worden sind.<sup>20</sup> Eine weitere Bestätigung für ihre Annahmen finden die Empiriker in der empirisch nicht nachweisbaren Korrelation zwischen der Höhe der Mortalität und dem Ausmaß ihrer Reduktion. Wenn es ein Limit gäbe, so müsste ihrer Meinung nach der Rückgang der Mortalität in Ländern, die sich näher an der oberen Grenze für die Lebenserwartung bei Geburt befinden, niedriger sein als für Länder, die noch weiter von dieser Grenze entfernt sind. Daher werden Mortalitätsprognosen, die sowohl eine verlangsamte Reduktion als auch ein ultimatives Niveau für die Sterblichkeit annehmen, von den Empirikern als fehlerhaft bewertet, da sie nicht nur den zukünftigen Anstieg der Lebenserwartung bei Geburt, sondern auch die zukünftige Zahl der Älteren unterschätzen.

Wissenschaftler, die sich bei der Mortalitätsprognose auf evolutionsbiologische Argumente stützen, schätzen die (unendliche) Extrapolation vergangener Mortalitätsentwicklungen in die Zukunft als mathematisch korrekt, aber als (evolutions)biologisch unplausibel ein [204]. Sie gehen davon aus, dass Unsterblichkeit nicht im Genom des Menschen angelegt sei. Vielmehr sei die Vergreisung des Menschen ein Nebenprodukt der Evolution, die primär Gene zur Gesundheit und Vitalität zur Erreichung der Geschlechtsreife, der Reproduktion und für die Aufzucht des Nachwuchses in einem natürlichen Selektionsprozess herausgebildet habe.<sup>21</sup> Folglich sei die Vergreisung von der Evolution nicht intendiert, sodass die Lebenserwartung bei Geburt durch Ausbleiben tödlicher Bedrohungen aus der Umwelt und durch medizinische Eingriffe in den menschlichen Alterungsprozess zwar steigen aber nicht unendlich werden könne [49, 203]. Deshalb gehen die Vertreter dieser Perspektive bei einer Mortalitätsprognose davon aus, dass die zukünftige Entwicklung der Mortalität von der vergangenen und gegenwärtigen Entwicklung abweicht. Sie untermauern diese Annahme,

---

<sup>20</sup>So listen Oeppen und Vaupel [201] zahlreiche in der Vergangenheit aufgestellte Limits für die Lebenserwartung bei Geburt auf — angefangen 1921 von Dublin bis in die Gegenwart von Olshansky et al. —, die bereits zum Zeitpunkt der Veröffentlichung oder einige Jahre danach überschritten worden sind.

<sup>21</sup>In diesem Zusammenhang ist auch die Theorie der antagonistischen Pleiotropie von Williams [311] zu nennen, die u. a. besagt, dass schädliche Auswirkungen von Genen nach der Reproduktionsphase nicht selektiert werden können, insbesondere wenn diese Gene vor und in der reproduktiven Phase (für die Fortpflanzung) vorteilhaft sind.

indem sie die unterschiedlichen altersspezifischen Mortalitätsreduktionen im Verlauf des 20. Jahrhunderts anführen: Während der Anstieg der Lebenserwartung bei Geburt in der ersten Hälfte hauptsächlich auf den Rückgang der Säuglings- und Müttersterblichkeit zurückzuführen ist, trägt in der zweiten Hälfte hauptsächlich die zunehmende Bedeutung degenerativer Krankheiten im Alter dazu bei. Die zukünftige Entwicklung der Mortalität kann insofern von der vergangenen und gegenwärtigen Entwicklung abweichen, als dass tatsächlich Möglichkeiten gefunden werden, in den menschlichen Alterungsprozess einzugreifen. Jedoch stehen sie einem entscheidenden Fortschritt in der Gentechnik zur wesentlichen Erhöhung der Lebenserwartung bei Geburt skeptisch gegenüber [206].

Vallin und Meslé [286] zeigen 2009 in einer erneuten umfassenden Untersuchung (unter Einschluss verschiedener Datenquellen), dass die maximale Lebenserwartung bei Geburt seit 1750 einem stückweise linearen Trend mit jeweils unterschiedlich hohen Anstiegen anstatt einem einheitlichen linearen Anstieg (wie ihn Oeppen et al. [201] beobachtet haben) folgt. Der Trend eines jeden Abschnitts kann mit medizinischen, sozialen und ökonomischen Entwicklungen begründet werden. Zudem wird auch hier die zeitliche Verschiebung des Beitrags bestimmter Alter zur Erhöhung der Lebenserwartung bei Geburt deutlich. Demnach findet heute der größte Mortalitätsrückgang in den hohen Altersstufen statt, jedoch hat dieser Fortschritt einen weitaus geringeren Effekt auf die Lebenserwartung bei Geburt als ein (ebenso großer absoluter) Mortalitätsrückgang in jungen Altern.<sup>22</sup> Folglich sprechen sich die Autoren für einen höchstwahrscheinlich verlangsamten Anstieg der Lebenserwartung bei Geburt in der Zukunft aus, wenn nicht eine neue, schwer vorstellbare, bahnbrechende Entwicklung eine Beschleunigung bewirke. Demnach sei die bloße Extrapolation des historischen Langzeittrends zu optimistisch.

Eine gesonderte Stellung nimmt die Diskussion über die Existenz einer oberen Grenze für die Lebenserwartung bei Geburt ein. Vertreter beider Positionen sind sich über die Unsicherheit der zukünftigen Entwicklung der Mortalität bewusst [124] und ziehen unterschiedliche Argumente zur Generierung der Annahmen für die zukünftige Lebenserwartung

---

<sup>22</sup>Vaupel [294] zeigt in diesem Zusammenhang, dass jung versterbende Personen zwar mehr Lebensjahre verlieren als ältere versterbende Personen. Gleichwohl kann aber eine Reduktion der Sterbefälle um einen bestimmten Prozentsatz in der Summe mehr Lebensjahre bei Älteren *sparen* als bei Jüngeren, da in höheren Altern deutlich mehr Sterbefälle stattfinden als in jüngeren Altern. Insofern sollte die Bedeutung des Sterblichkeitsrückgangs in den hohen Altersstufen bei der Erhöhung der Lebenserwartung (bei Geburt) nicht unterschätzt werden.

bei Geburt heran. Die Vertreter der evolutionsbiologischen Perspektive sprechen sich für eine obere Grenze aus, die durch einen verlangsamten Anstieg in der (nahen) Zukunft erreicht werden könne. Dies begründen sie mit den bereits oben aufgeführten theoretischen Überlegungen u. a. zur Evolution. Daraus schließen sie, dass die Lebenserwartung bei Geburt einen bestimmten Wert nicht überschreite, sofern keine lebensverlängernden Eingriffe in den menschlichen Alterungsprozess durch medizinische, gentechnische Innovationen durchgeführt werden können [49, 317]. Im Gegensatz dazu zeigen die Empiriker anhand empirischer Daten, dass bisher keine obere Grenze für die Lebenserwartung bei Geburt erreicht worden ist. So sinken in etlichen Industrieländern die Sterberaten stetig und in den hohen Altern findet seit 1970 sogar ein beschleunigender Rückgang der Sterblichkeit statt. Auch die empirisch nicht nachweisbare Korrelation zwischen der Höhe der Mortalitätsreduktion und dem Mortalitätsniveau wird als ein weiteres Indiz dafür angesehen, dass sich der gegenwärtige Wert der Lebenserwartung bei Geburt nicht in der Nähe einer oberen Grenze befindet [124, 224, 296, 297, 299, 316]. Darüber hinaus steigt die maximale individuelle Lebensdauer in etlichen Industrieländern seit über 100 Jahren und in letzter Zeit sogar verstärkt (linear) an [312, 313, 316]. Darüber hinaus wird auch dem Lebensstil mit der vermehrt bewussten Vermeidung gesundheitsgefährdender Risikofaktoren ein bedeutender Einfluss auf die Erhöhung der zukünftigen Lebenserwartung bei Geburt zugeschrieben [183]. Daraus schließen die Empiriker generell, dass keine obere Grenze für die Lebenserwartung bei Geburt in naher Zukunft erreicht würde. Die Annahme, dass es keine (naheliegende) obere Grenze gibt, hieße jedoch nicht, dass die Sterblichkeit abrupt auf 0 sinke und damit Unsterblichkeit vorherrsche. Vielmehr bedeute dies, dass die Lebenserwartung bei Geburt weiter ansteigen könne [312]. Generell könne aber auch aufgrund der geringen Datenbasis insbesondere in den hohen Altern solch eine obere Grenze nicht gänzlich ausgeschlossen werden. Wenn es demnach ein Limit gäbe, könne sich dieses mit der Zeit durch z. B. medizinische, gentechnische Innovationen oder auch durch bestimmte Verhaltensänderungen verändern bzw. in ein jeweils noch höheres Alter verschieben [316, 298].

Zusammenfassend kann festgehalten werden, dass sich die Positionen letztlich gar nicht so stark voneinander unterscheiden. Während die Empiriker davon überzeugt sind, dass die Lebenserwartung bei Geburt durch medizinische Innovationen in der Zukunft wie in der Vergangenheit weiter ansteigen wird, schwächen die Vertreter der evolutionsbiologischen Perspektive diese Aussage ab, indem sie sagen, dass wenn medizinische Innovationen stattfinden, die Lebenserwartung bei Geburt weiter steigen kann. Gibt es jedoch keine medizinischen

Innovationen, so erreicht die Lebenserwartung bei Geburt aus evolutionsbiologischer Perspektive eine (naheliegende) obere Grenze, die aus Sicht der Empiriker noch lange nicht erreicht ist. Eine wissenschaftlich fundierte und sachkundige Diskussion zu diesem Thema ist allein mit der Demographie nicht zu bewerkstelligen. Die Demographie kann lediglich statistische Maßzahlen zur Mortalität und damit zur Langlebigkeit in der Vergangenheit und Gegenwart liefern. Darüber hinaus ist die rein demographische Diskussion über die zukünftige Entwicklung der Langlebigkeit eher spekulativ, da sie zusätzliche Informationen u. a. aus der Biologie, Systembiologie, Medizin und Gerontologie bedarf, die selbst in der Gegenwart zu diesem Thema noch auf dem Stand der Entwicklung und Forschung sind [65, 203, 204]. Die Diskussion über die Langlebigkeit der Menschen hat insbesondere mit der Aufdeckung des menschlichen Genoms an Fahrt gewonnen. Jedoch gestalten sich die Fortschritte in der Biologie und Systembiologie schleppend, sodass die seit der Entdeckung gemachten Erwartungen an die Hilfe in der Medizin bislang noch nicht erfüllt werden konnten. So reicht die Kenntnis der DNS (Desoxyribonukleinsäure) und ihrer einzelnen Bestandteile wie z. B. Gene und die mit ihnen zu generierenden Proteine nicht aus. Vielmehr wird ein umfassendes Verständnis des komplexen lebenden Systems Mensch von den Genen bis zum gesamten Organismus benötigt, um fundierte Mutmaßungen über die Langlebigkeit anstellen zu können [199]. Wenn demnach keine zusätzlichen gesicherten Informationen aus relevanten Fachbereichen vorliegen, bleibt bei einer Sterblichkeitsprognose lediglich der Zugriff auf die empirisch messbare Sterblichkeitsentwicklung in der Vergangenheit und Gegenwart. Bereits Pritchett ([222], S. 278) hat 1891 festgestellt, dass dadurch die Genauigkeit einer Prognose nicht unbedingt geschmälert werden muss, wenn nur die (zum Zeitpunkt der Prognoseerstellung nicht ganz verstandenen) Bedingungen der Sterblichkeitsentwicklung weiterhin anhalten.

Für die Generierung von Mortalitätsannahmen sind dieser Diskussion zwei Alternativen zu entnehmen: Einerseits die Extrapolation vergangener Trends in die Zukunft und andererseits die Abschwächung vergangener Trends in der näheren Zukunft bis zu einem ultimativen Niveau.

### **Mortalitätsprognose mit Berücksichtigung von Todesursachen**

Wie die Diskussion über die Langlebigkeit des Menschen (vgl. Abschnitt 5.2.3 ab Seite 98) bereits gezeigt hat, ist ab der zweiten Hälfte des 20. Jahrhunderts eine zunehmende demographische Alterung in etlichen Industrieländern beobachtet worden. Während anfangs die Reduktion der Sterblichkeit in allen Altern beobachtet werden konnte, ist

mit der Zeit die Frage nach den Ursachen für diese Entwicklung aufgeworfen worden. Dabei sollte die Disaggregation der gesamten Sterblichkeitsentwicklung nach dem Alter, dem Geschlecht und nach einzelnen Todesursachen einen zusätzlichen Erkenntnisgewinn erbringen [11, 64, 97, 186, 220, 275].<sup>23</sup>

Die vergangene Entwicklung der Gesamtsterberate ist zwar relativ stabil und regelmäßig, jedoch ist sie zu verschiedenen Zeiten von unterschiedlichen Haupttodesursachen bestimmt bzw. dominiert worden [315]. Sofern der jeweils aus unterschiedlichen Entwicklungen zusammengesetzte Haupttrend der Sterblichkeit nicht von seiner vergangenen Entwicklung abweicht, ist eine Prognose der Gesamtsterblichkeit angebracht. Sofern jedoch die heterogene Ursachenstruktur zu einer Abweichung der Gesamtsterblichkeit von ihrem bisherigen Trend führt, könnte die Prognose der Gesamtsterberate als Summe der todesursachenspezifischen Sterberaten genauere Prognoseergebnisse generieren.

Mit der Berücksichtigung der Todesursachenstruktur auf die Entwicklung der Gesamtsterblichkeit können sowohl direkt die Effekte einzelner Todesursachen auf die Gesamtsterblichkeit als auch indirekt die Auswirkungen einzelner Veränderungen im Gesundheitsverhalten (oder auch in der Wirksamkeit medizinischer Therapien) auf die Entwicklung der Sterblichkeit einer Todesursache ermittelt werden [97, 275]. Beispielhaft sei hier an die Erhöhung der Lungenkrebssterblichkeit bei den Frauen durch den vermehrten Tabakkonsum in letzter Zeit oder auch an den Rückgang des Gebärmutterhalskrebses durch vermehrte Prävention erinnert [51, 312]. Außerdem sollten die Ergebnisse der Prognose der Gesamtsterblichkeit nicht mehr durch unberücksichtigte Heterogenität verzerrt werden, da nun die einzelnen Ursachenkomponenten in Form der Todesursachen separat betrachtet und zu einer Prognose der Gesamtsterblichkeit addiert werden.

Jedoch können auch einige Probleme bei der Berücksichtigung von Todesursachen in einer Prognose der Gesamtsterblichkeit auftreten. So ist die vergangene Entwicklung der todesursachenspezifischen Sterberaten weitaus weniger stabil und regelmäßig als die der Gesamtsterberate [314, 315]. Darüber hinaus erschwert die regelmäßige Überarbeitung des international einheitlichen Systems zur Klassifizierung der Todesursachen und Krankhei-

---

<sup>23</sup>Thompson und Whelpton [281] oder auch Dublin und Lotka [76] haben bereits in der ersten Hälfte des 20. Jahrhunderts herausgestellt, dass bei einer Prognose der Sterblichkeit auch Krankheiten und Todesursachen berücksichtigt werden sollten.

ten<sup>24</sup> die konsistente Analyse der Sterblichkeit an einer Todesursache über die Zeit, da mit der Herausgabe neuer Revisionen Krankheiten nicht mehr aufgeführt werden, durch Ausdifferenzierung anders kodiert oder neu hinzukommen können [64, 220, 275]. Außerdem können die Todesursachendaten nicht immer fehlerfrei erhoben werden. Im medizinischen Teil eines Totenscheins muss (in der Bundesrepublik) vom Arzt eine Kausalkette vom Grundleiden bis zur unmittelbar zum Tode führenden Krankheit angegeben werden. Erste Erfassungsunreinheiten können demnach durch ärztliche Fehldiagnosen [51] entstehen; zumal die ärztliche Diagnose bei Sterbefällen im höheren Altern durch relativ häufig auftretende Multimorbidität erschwert wird, da immer nur ein Grundleiden angegeben werden kann. Zusätzliche Erfassungsprobleme können außerdem durch Fehlkodierungen des Grundleidens in den zuständigen Statistikämtern entstehen.<sup>25</sup> Um diesen Problemen zumindest teilweise zu entgehen, sollten nur Haupttodesursachengruppen wie z. B. Erkrankungen des Kreislaufsystems oder bösartige Neubildungen verwendet werden. Aber auch unter der Annahme, alle Todesursachen würden amtlich richtig erfasst werden, besteht immer noch das Problem der so genannten konkurrierenden Risiken. Darunter ist eine Verzerrung der Prognoseergebnisse der Gesamtsterblichkeit durch die unabhängige Prognose der einzelnen todesursachenspezifischen Sterberaten zu verstehen, obwohl die Todesursachen in der Realität weder unabhängig sind noch unabhängig voneinander betrachtet werden können [11, 33, 315, 319].

Generell sollte ein ideales Modell zur Prognose der Gesamtsterblichkeit alle relevanten Einflussfaktoren berücksichtigen. Sind jedoch nicht alle relevanten Einflussfaktoren in einem entsprechenden Mortalitätsprognosemodell enthalten, können möglicherweise auch nicht alle Trendentwicklungen erkannt und in der Prognose berücksichtigt werden. Eine Möglichkeit besteht nun darin, die Entwicklung der Gesamtsterblichkeit in ihre Teilentwicklungen zu zerlegen. Dazu kann die Gesamtsterblichkeit als Summe der Sterblichkeit einzelner Todesursachen prognostiziert werden. Verschiedene (bereits oben besprochene) Methoden können zur Prognose der todesursachenspezifischen Sterberaten verwendet werden.

Häufig kommt bei der Prognose der todesursachenspezifischen Sterberaten die Extrapolation zum Einsatz. Allerdings ist eine Extrapolation der todesursachenspezifischen Sterberaten nur geeignet, wenn die in der Vergangenheit aufgetretenen Trends auch in der Zukunft weiter anhalten [315]. Folglich muss die extrapolierte Gesamtsterblichkeit nach

---

<sup>24</sup>ICD - International Classification of Diseases and Related Health Problems

<sup>25</sup>So werden in der aktuellen ICD 10 bis zu 13.000 Positionen unterschieden [245].



Todesursachen nicht zwingend genauer sein als die Extrapolation der Gesamtsterblichkeit allein. So können zum Zeitpunkt der Prognoseerstellung unerwartete Trendumkehrungen mit einer reinen Extrapolation in der Vergangenheit beobachteter Trends nur ungenügend bzw. gar nicht antizipiert werden. Um solche zukünftigen Trendumkehrungen antizipieren zu können, müssen nicht nur vergangene empirische Entwicklungen, sondern z. B. auch Annahmen zu relevantem Verhalten und medizinischen Innovationen und Therapien in eine Prognose für einzelne Todesursachen mit einbezogen werden [315]. Finden keine unerwarteten Trendumkehrungen statt, ist die Extrapolation der Sterblichkeit nach Todesursachen ähnlich genau wie die der Gesamtsterblichkeit allein [186].

Die Untersuchung von Caselli [52] zeigt, dass die Extrapolation der Gesamtsterberate als Summe der todesursachenspezifischen Sterberaten keinen entscheidenden Vorteil gegenüber der Prognose der Gesamtsterberate allein erbringt. Zudem ist die Prognose der Gesamtsterberate nach Todesursachen meist höher bzw. pessimistischer als die Prognose der Gesamtsterberate allein [52, 64, 314, 315]. So wird der anteilmäßig sinkende positive Effekt in der Sterblichkeit überdurchschnittlich sinkender Todesursachen (wie z. B. Unfälle bei den Jungen) durch den wachsenden negativen Effekt in der Sterblichkeit unterdurchschnittlich sinkender oder sogar steigender Todesursachen (wie z. B. degenerative Krankheiten) überlagert, sodass daraus häufig eine höhere kumulierte Gesamtsterblichkeit resultiert [314].

Um die Extrapolation der Gesamtsterblichkeit nach Todesursachen zu verbessern, schlägt Crimmins [64] vor, die Einflussfaktoren einer jeden Todesursache zu rekapitulieren und sich dementsprechend Trends von Basisperioden unterschiedlicher Länge herauszusuchen, die diese Annahmen am besten widerspiegeln. Häufig wird dabei auf die zuletzt beobachteten Trends zurückgegriffen werden.

Eine weitere Möglichkeit zur Prognose der Gesamtsterblichkeit nach Todesursachen besteht in der Verwendung eines Mortalitätsgesetzes. Aber auch hier wird wie bei der Extrapolation die vergangene Entwicklung der Modellparameter mit einem geeigneten Zeitreihenmodell fortgeschrieben, ohne zusätzliche Informationen zu berücksichtigen [186]. Demnach stellen Mortalitätsgesetze keine Verbesserung bei der Prognose der Gesamtsterblichkeit nach Todesursachen dar.

Grundsätzlich gibt Alho [5] zu bedenken, dass bei einer Prognose der Gesamtsterblichkeit

nach Todesursachen vermehrt Fehler durch Modellmisspezifikation und Missklassifikation auftreten und das Ergebnis verzerren können. Daneben ist die Prognose der Mortalität nach Todesursachen gerade in den höchsten Altern durch das vermehrte Auftreten von Multimorbidität und der dadurch erschwerten Diagnose des zum Tode geführten Grundleidens problematisch. Das ist um so gravierender für die gesamte Mortalitätsprognose als dass gerade in diesem Altersbereich das größte Potential für eine Mortalitätsreduktion steckt [275].

Zusammenfassend ist festzuhalten, dass die Prognose der Gesamtsterblichkeit nach Todesursachen nur dann genauere Ergebnisse als die Prognose der Gesamtsterblichkeit allein generiert, wenn Veränderungen in der Ursachenstruktur durch z.B. Trendumkehrungen auftreten und diese durch geeignete Modellparameter eingefangen werden können. Die bisher häufig angewendete Extrapolation der Gesamtsterblichkeit nach Todesursachen erbringt keine genaueren Ergebnisse als die Extrapolation der Gesamtsterblichkeit allein, da sie auch ausschließlich die in der Vergangenheit beobachteten Trends fortschreibt und keine zusätzlichen Informationen berücksichtigen kann. Anstatt dessen enthält die Prognose der Gesamtsterblichkeit nach Todesursachen zusätzlich Unsicherheit durch höhere Komplexität. Damit liefert sie weder einen methodischen noch einen inhaltlichen Vorteil gegenüber der Extrapolation der Gesamtsterblichkeit allein, wenn die Genauigkeit der Prognoseergebnisse das ausschlaggebende Kriterium ist. Sofern es bei einer Prognose aber auch um weitere Erkenntnisgewinne geht, also z.B. wie bedeutend bestimmte Todesursachen für die Entwicklung der Gesamtsterblichkeit sind oder wie wichtig die Transparenz der getroffenen Annahmen ist, so kann die Prognose der Gesamtsterblichkeit nach Todesursachen einen entscheidenden Beitrag liefern [64, 97, 186, 275]. Jedoch muss dann eine Methode gefunden werden, mit der zusätzliche Informationen aus z.B. allgemein anerkannten biologischen Theorien zum Alterungsprozess in der Zukunft berücksichtigt werden können, damit die Prognose der Sterblichkeit nach Todesursachen nicht nur genauere, sondern auch zusätzliche Erkenntnisse liefert [315].

Eine Mortalitätsprognose nach Todesursachen auf dem derzeitigen Entwicklungsstand ist (wenn überhaupt) eher für Kurz- als für Langzeitprognosen geeignet, da sie dann mit einer höheren Wahrscheinlichkeit von einem Fortdauern der jeweils bis dahin beobachteten Trends ausgehen kann [11].

Dem Unterabschnitt Mortalitätsprognosen mit Todesursachen sind noch ergänzend der An-

satz der Todesursachenelimination wie auch der Cause-Delay Ansatz zuzuordnen. Bei dem Ansatz der Todesursachenelimination geht es um die grundlegende Frage, wie hoch die Lebenserwartung bei Geburt wäre, wenn eine (oder mehrere) Todesursachen eliminiert werden würden. Problematisch erweist sich dabei die Schätzung der Sterbewahrscheinlichkeiten an den verbleibenden Todesursachen, da dazu die (zu verhindernden) Sterbefälle an der zu eliminierenden Todesursache auf die übrigen Todesursachen verteilt werden müssen.<sup>26</sup> Allerdings ist die Sterblichkeit nach Todesursachen nicht für jede einzelne Todesursache separat (bzw. isoliert) messbar, da die Todesursachen ausschließlich voneinander abhängig auftreten<sup>27</sup> [182, 191]. Der Cause-Delay Ansatz analysiert dagegen, wie sich eine Verschiebung der Sterbefälle in ein höheres Alter auf die Lebenserwartung bei Geburt auswirkt. Diese Herangehensweise ist im Vergleich zur Todesursachenelimination plausibler, da beispielsweise eine gesündere Lebensweise oder auch eine wirksamere medizinische Behandlung den Tod an einer Todesursache hinauszögern kann. Darüber hinaus müssen beim Cause-Delay Ansatz auch keine zu verhindernden Sterbefälle, die aufgrund einer Todesursachenelimination anfallen, auf übrige Todesursachen aufgeteilt werden, da sich nicht die Anzahl der Sterbefälle an den einzelnen Todesursachen, sondern lediglich der Zeitpunkt der Sterbefälle an den einzelnen Todesursachen verändert. Doch auch bei der Bestimmung der altersspezifischen Sterbewahrscheinlichkeiten nach (zu verschiebenden) Todesursachen tritt das Problem der konkurrierenden Risiken auf. Eine Möglichkeit zur Schätzung der Sterbewahrscheinlichkeiten in den Altern  $x$  an einer (zu verschiebenden) Todesursache besteht darin, jeweils die Werte der Sterbewahrscheinlichkeiten aus vorherigen Altern  $x - n$  heranzuziehen. Die Verwendung der Sterbewahrscheinlichkeitswerte jüngerer Alter aus derselben Bevölkerung hat den Vorteil, dass die mortalitätsbedingenden Faktoren unverändert bleiben [182, 202, 256].

### 5.2.4. Sonstige Ansätze

Neben den extrapolativen, interpolativen und erklärenden bzw. multikausalen Ansätzen gibt es sonstige Ansätze, die in letzter Zeit im Bereich der Mortalitätsprognose an Bedeutung gewonnen haben. Dazu zählen z. B. bayesianische Ansätze oder auch Ansätze, die Tempoeffekte in einer Mortalitätsprognose berücksichtigen.

---

<sup>26</sup>Würden die zu verhindernden Sterbefälle nicht auf andere Todesursachen aufgeteilt werden, würde für diese Personen Unsterblichkeit angenommen werden.

<sup>27</sup>Hierbei handelt es sich um das Problem der konkurrierenden Risiken [191, 319].

### Bayesianische Ansätze

Es gibt eine Vielzahl bayesianischer Ansätze, die erst in den letzten Jahren im Bereich der Mortalitätsprognose Beachtung finden.<sup>28</sup> Bayesianische Ansätze können einen Mortalitätsparameter in Abhängigkeit diverser Erklärungsvariablen (bzw. Modellparameter) unter Berücksichtigung der Prognoseunsicherheit flexibel prognostizieren. So können bei der Generierung der Annahmen für die Modellparameter quantitative und qualitative Informationen unterschiedlicher Quellen in der jeweiligen *prior distribution* variabel vereint werden. Dabei können die Informationen aus den verfügbaren Daten (*likelihood*) um zusätzliche subjektive Expertenmeinung (*prior*) ergänzt werden. Nach Einsetzen der Annahmen in die Modellgleichung werden die zukünftigen Werte des zu prognostizierenden Mortalitätsparameters in einer *posterior distribution* ausgegeben (vgl. Formel 5.10). Die Prognoseunsicherheit wird sowohl über die Wahrscheinlichkeitsverteilung der Modellparameter als auch über die Wahrscheinlichkeitsverteilung der Ergebnisse widerspiegelt.

$$Posterior \propto Likelihood \cdot Prior \quad (5.10)$$

Giroso und King [96, 97] oder auch Pedroza [212] nutzen beispielsweise bayesianische Ansätze zur Prognose der altersspezifischen Sterblichkeit.

Bayesianische Ansätze schaffen prinzipiell eine gelungene Symbiose aus quantitativen und zusätzlichen qualitativen Informationen unter Berücksichtigung der Prognoseunsicherheit in entsprechenden Wahrscheinlichkeitsverteilungen für die Modellparameter und die jeweilige Ergebnisgröße. Damit können quasi alle verfügbaren Informationen und subjektiven Annahmen flexibel in einer Prognose integriert werden. Allerdings sind die bayesianischen Ansätze methodisch sehr anspruchsvoll und mit einem relativ hohen Datenaufwand verbunden. Außerdem versuchen auch diese Ansätze alle Informationen in eine Funktionsgleichung zu integrieren, wodurch die Vielfältigkeit eingeschränkt werden kann.

### Beachtung von Tempoeffekten in der Mortalitätsprognose

In der neueren Zeit sind Überlegungen zu Tempoeffekten in der Mortalität aufgekommen [29, 177]. Der grundlegenden Idee nach werde die jährlich auszuweisende Lebenserwartung bei Geburt über- oder unterschätzt, wenn die Sterblichkeit gleichzeitig von mehreren Kohorten

---

<sup>28</sup>Eine Einführung zu bayesianischen Ansätzen ist z. B. bei Gelman [93], Carlin und Louis [47] oder auch bei Lynch [179] zu finden.

in ein höheres oder früheres Alter verschoben wird. Solche Verzerrungseffekte sollten bei einer Mortalitätsprognose durch die Berechnung einer tempo-adjustierten Lebenserwartung bei Geburt berücksichtigt werden. Allerdings sind diese beschriebenen Verzerrungseffekte in der Mortalität umstritten [101]. So gibt es sogar gegensätzliche Überlegungen, die besagen, dass die Sterblichkeit in einem Jahr bei Verlagerung ins höhere Alter überschätzt werde, da die einzelnen altersspezifischen Sterberaten nicht nur von der aktuellen Periode, sondern auch von verschiedenen Bedingungen vergangener Perioden beeinflusst werden. Demnach können bei der Durchführung einer Mortalitätsprognose Überlegungen angestellt werden, inwiefern Tempoeffekte und Auswirkungen der Heterogenität in der Sterblichkeit [295] auf zusammenfassende Mortalitätsparameter wie die Lebenserwartung bei Geburt berücksichtigt werden sollen.

### 5.3. Schlussfolgerung — Vorschläge zur Generierung von Mortalitätsannahmen für eine Prognose mit dem PPPM

Wie in diesem Kapitel gezeigt worden ist, gibt es eine Vielzahl von Ansätzen zur Mortalitätsprognose. Ein wesentliches Unterscheidungsmerkmal stellt die Art der dabei zu nutzenden Informationen dar. Schreiben pure Extrapolationsansätze ausschließlich die historische Trendentwicklung der (willkürlich zu bestimmenden) Basisperiode fort, nutzen interpolative Ansätze, multikausale Ansätze und bayesianische Ansätze auch zunehmend weitere, qualitative (bzw. subjektive) Informationen zur Mortalitätsprognose.<sup>29</sup> Bei interpolativen Ansätzen kann zumindest die Zielverteilung subjektiv bestimmt werden, wohingegen der Entwicklungspfad von der Start- zur Zielverteilung über ein rein mathematisch interpolatives Verfahren bestimmt werden kann. Dagegen können bei multikausalen Ansätzen vielfältige theoretische Überlegungen und Erkenntnisse relevanter Wissenschaftsbereiche subjektiv bei einer Mortalitätsprognose integriert werden, indem z. B. die Fortführung eines (abgeschwächten bzw. verstärkten) in der Vergangenheit observierten Trends oder auch eine Trendumkehr mit einem geeigneten parametrischen Verfahren angenommen werden kann. Zudem kann auch das Aggregationsniveau des zu prognostizierenden Parameters verändert werden, um die Prognosekraft durch zusätzliche Informationen zu erhöhen. So

---

<sup>29</sup>Die in diesem Kapitel aufgestellte Systematik der zu unterscheidenden Ansätze zur Mortalitätsprognose orientiert sich an dem Nutzungsgrad zusätzlicher subjektiver Informationen. Infolgedessen gibt es keine klare Trennung zwischen den rein mathematisch-statistischen Verfahren, sodass z. B. eine Extrapolation in allen vier Kategorien wiedergefunden werden kann.

kann beispielsweise das Aggregationsniveau der zu prognostizierenden Sterberate reduziert werden, um todesursachenspezifische Sterberaten unter Einsatz zusätzlicher, todesursachen-betreffender Informationen zu prognostizieren. Daneben können quantitative und qualitative Informationen mit bayesianischen Ansätzen flexibel (und subjektiv) mit einer entsprechenden Wahrscheinlichkeitsverteilung berücksichtigt werden.

Die bestehenden Ansätze zur Mortalitätsprognose können nur sinnvolle und genaue Ergebnisse liefern, wenn die in der Vergangenheit und Gegenwart beobachteten Erfahrungen auch in der Zukunft gelten, da unerwartete Entwicklungen nur sehr schwer eingefangen werden können [97]. Jedoch versuchen Wissenschaftler der Medizin oder Gentechnik in den Alterungsprozess des Menschen einzugreifen, um diesen mit möglichst gesunden Lebensjahren zu verlängern. Folglich handelt es sich hierbei um zukünftige Ereignisse, die zwar erwartet werden können, aber noch nicht in der Vergangenheit beobachtet werden konnten. Solche Informationen oder Erwartungen können nur über Expertenmeinung Eingang in eine Prognose finden. Deshalb sollten Ansätze zur Mortalitätsprognose nicht nur die quantitative Entwicklung des zu prognostizierenden Mortalitätsparameters, sondern zudem noch weitere, qualitative Informationen aus diversen relevanten wissenschaftlichen Fachbereichen wie der Medizin und Gentechnik berücksichtigen.

Demnach ist der Einsatz aller verfügbaren Informationen auf den ersten Blick zwingend notwendig, um eine genaue Mortalitätsprognose zu erstellen. Allerdings hat sich gezeigt, dass das nicht immer zutreffen muss. Zumal die Informationen über mortalitätsbedingende Faktoren oftmals noch ungesichert sind, sodass ein eindeutiger Effekt (in Richtung und Ausmaß) auf die Sterblichkeit in der Zukunft nur sehr schwer korrekt ermittelt werden kann.

Eine Schlussfolgerung aus dieser Diskussion besteht darin, sich bei einer Mortalitätsprognose vornehmlich auf die gesicherten (demographischen) Daten zu stützen, die dann mit weiteren Informationen (korrigierend) abgeändert werden können. So ist die Extrapolation der (vergangenen) Sterblichkeitsentwicklung zu empfehlen, wenn dabei die Wahl (der Basisperiode wie auch) des zu verwendenden parametrischen Modells subjektiv unter Einschluss zusätzlicher Informationsquellen wie z.B. beim multikausalen Ansatz zur Langlebigkeit getroffen wird. Es steht zwar fest, dass pure Extrapolationsmethoden eine relativ sichere Datenbasis nutzen, allerdings ist ihre Annahme gleichbleibender stabiler Bedingungen nur über einen relativ kurzen Projektionshorizont hinreichend plausibel. Demzufolge sollten mit

zunehmender Länge des Projektionshorizonts zusätzliche Informationen bei einer Prognose der Sterblichkeit zunehmend genutzt werden, um potentielle Trendumkehrungen einfangen zu können. Allerdings sollte dabei stets die Sicherheit des zusätzlichen Informationsgehalts mit dem entsprechenden Informationsgewicht in der Mortalitätsprognose korrespondieren.

Bei der beispielhaften Anwendung des PPPM (vgl. Kapitel 12) werden deshalb zur Generierung der Mortalitätsannahmen Extrapolationsmethoden in Verbindung mit Expertenmeinung bzw. unter Einschluss zusätzlicher Informationen genutzt werden.

## 6. Generierung von Fertilitätsannahmen

Die Prognose der relativ instabilen Fertilitätsentwicklung ist komplex und unsicher, da sie von für einen Prognostiker oftmals schwer einschätzbaren individuellen und sozio-ökonomischen Faktoren abhängt.

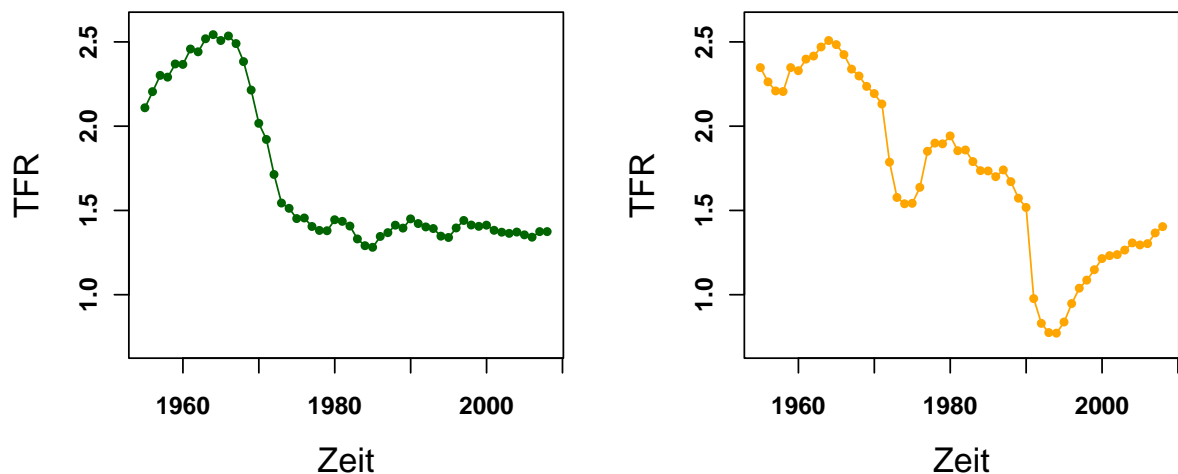


Abbildung 6.1.: Entwicklung der TFR in den alten (links) und neuen (rechts) Bundesländern von 1955 bis 2008; Datenquelle: Statistisches Bundesamt

In der Abbildung 6.1 ist zu erkennen, dass die Entwicklung der TFR<sup>1</sup> in den alten und neuen Ländern in der Zeit von 1955 bis 2008 relativ stark schwankt. In den alten Ländern folgt in dieser Zeit dem Hoch der TFR von 2,5425 im Jahr 1964 ein rapider Abfall bis zum Ende der 1970er Jahre. Seitdem schwankt die TFR in den alten Ländern um den Wert von ungefähr 1,38. In den neuen Ländern zeigt sich hingegen ein anderes Bild. Auch hier folgt

---

<sup>1</sup>Die *Total Fertility Rate*, kurz die TFR, bildet sich für ein Jahr  $t$  über die Summe aller altersspezifischen Fertilitätsraten (vgl. Formel 9.1 auf Seite 174) im Jahr  $t$ . Damit gibt die TFR eines Jahres  $t$  die durchschnittliche Kinderzahl einer Frau in einem Jahr  $t$  unter der Bedingung an, dass diese Frau die gesamte reproduktive Phase überlebt und dass für sie in jedem Alter die angenommenen Fertilitätsraten gelten [221].



zwar dem Hoch von 2,50576 im Jahr 1964 ein rapider Rückgang der TFR bis zur Mitte der 1970er Jahre. Allerdings steigt die TFR anschließend u. a. aufgrund der Einführung pronatalistischer Familienmaßnahmen<sup>2</sup> bis 1980 auf einen Wert von 1,9418 an, um dann bis 1990 leicht und dann aufgrund der Wende bis 1994 fast sprunghaft auf einen Tiefstwert von 0,7722 abzusinken. Seitdem steigt die TFR in den neuen Ländern kontinuierlich an, sodass sie im Jahr 2008 einen Wert von 1,404<sup>3</sup> erreicht hat.

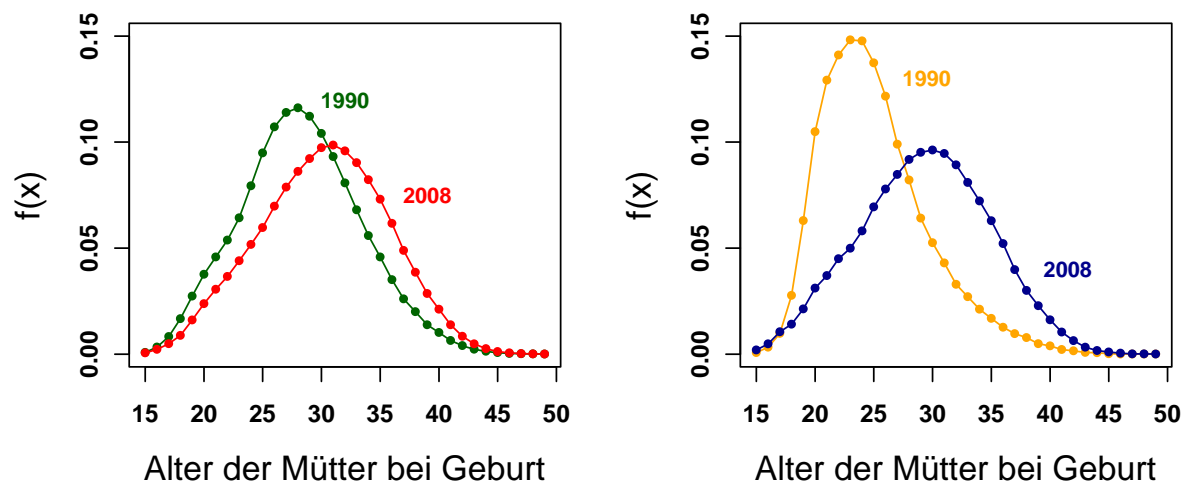


Abbildung 6.2.: Altersspezifische Fertilitätsraten in den alten (links) und neuen (rechts) Bundesländern 1990 und 2008; Datenquelle: Statistisches Bundesamt

Daneben zeigen die Abbildungen 6.2 und 12.6 (auf der Seite 254), dass die Fertilität in den alten und neuen Ländern wie auch in Deutschland eine relativ stabile Altersverteilung aufweist, die einer Glockenkurve ähnelt. Allerdings wird die Verteilung der altersspezifischen Fertilitätsraten einerseits immer flacher und andererseits verschiebt sie sich zunehmend nach rechts. Diese beiden Effekte werden auch als Quantum- und Timingeffekt bezeichnet, die (in dieser Ausprägung) einerseits den Rückgang der Fertilität und andererseits das Aufschieben der Geburt von Kindern in ein höheres Alter im Lebensverlauf beschreiben. Die Abbildung 6.2 zeigt zudem die Annäherung der Fertilität zwischen den neuen und alten Ländern seit

<sup>2</sup>So ist der Geburtenanstieg in der zweiten Hälfte der 1970er nach Huinink et al. [115] mitunter der Einführung eines *Babyjahres* 1976 zuzuschreiben, nach der erwerbstätige Mütter ab der Geburt eines zweiten Kindes ein Jahr vollen Lohnausgleich erhalten.

<sup>3</sup>Der Wert der TFR in der Höhe von 1,404 gilt für die neuen Länder ohne Berlin nach der Geburtsjahrmethode. Wird der Wert der TFR für die neuen Länder mit Berlin nach der Geburtsjahrmethode ausgegeben, so liegt er bei 1,373.

1990.

Da die Fertilität in ihrer Höhe relativ stark vom Alter der Frauen abhängt, sollte ein Parameter zur Fertilitätsprognose verwendet werden, der sensibel gegenüber Veränderungen der Fertilität in einzelnen Altern ist. Demzufolge stellen die altersspezifischen Fertilitätsraten (vgl. Abschnitt 9.2.3 ab Seite 173) einen geeigneten Parameter zur Prognose der Fertilität dar. Die TFR eignet sich indes weniger für eine genaue Fertilitätsprognose, da verschiedene Verteilungen der altersspezifischen Fertilitätsraten denselben Wert für die TFR erzielen können.

Neben dem Alter variiert die Fertilität auch nach der Parität einer Frau. Allerdings ist die Analyse paritätsspezifischer Fertilitätsraten in Deutschland mit amtlichen Daten unmöglich, da zufolge des Bevölkerungsstatistikgesetzes [42] nur die Geburtenfolge von Kindern einer bestehenden Ehe erfasst wird. Deshalb sollte für paritätsspezifische Fertilitätsanalysen auf andere Datenquellen zurückgegriffen werden, die Personen nach der Kinderzahl — unabhängig von deren (II)Legitimität — befragen.<sup>4</sup>

Darüber hinaus können statistische Maßzahlen zur Messung der Fertilität sowohl im Querschnitt für eine Periode als auch im Längsschnitt für eine Geburtskohorte erstellt werden. Da Periodenmaße die Fertilität einer fiktiven Kohorte bestimmen, können sie durch echte Kohorteneffekte verzerrt werden. Ein Kohorteneffekt kann beispielsweise darin bestehen, dass Geburten durch ein sozial und ökonomisch einschneidendes Ereignis (wie z. B. die Wende in den neuen Ländern in Deutschland) in ein höheres Alter verschoben werden, sodass die TFR, ein Periodenmaß, sinkt, obwohl sich die tatsächliche Fertilitätshöhe nicht verändert. Derlei Verzerrungseffekte können bei Kohortenmaßen, wie z. B. der CFR,<sup>5</sup> nicht auftreten, da sie die Fertilität echter Geburtskohorten über die gesamte reproduktive Altersspanne bestimmen, wodurch weder ein Aufschub noch eine Vorverlegung von Geburten im Lebensverlauf die Höhe der CFR beeinflussen kann.

---

<sup>4</sup>So nutzen z. B. Kreyenfeld et al. [146] den Mikrozensus, um Aussagen über Fertilitätsbiographien zu treffen.

<sup>5</sup>Die *Cohort Fertility Rate*, CFR, gibt (analog zur TFR) die durchschnittliche Kinderzahl einer Frau einer echten Geburtskohorte an. Sie wird demzufolge aus der Summe der altersspezifischen Fertilitätsraten einer Geburtskohorte berechnet.

## 6.1. Überblick über verschiedene Ansätze zur Fertilitätsprognose

Es gibt eine Vielzahl bestehender Ansätze zur Prognose der Fertilität, die sich hinsichtlich des zu prognostizierenden Fertilitätsparameters, der Prognosemethode wie auch der zu verwendenden Informationen (wie beispielsweise der historischen Entwicklung oder auch Theorien zum Fertilitätsverhalten) unterscheiden.

In diesem Abschnitt kann lediglich ein Einblick in die Vielfalt der bestehenden Ansätze zur Fertilitätsprognose gegeben werden.<sup>6</sup> Die Einteilung der verschiedenen Ansätze orientiert sich dabei an der entwickelten Systematik für die Ansätze zur Mortalitätsprognose (vgl. Abschnitt 5.2 ab Seite 82), sodass extrapolative von interpolativen, theoriebasierten und multikausalen Ansätzen zur Fertilitätsprognose unterschieden werden können.

### 6.1.1. Extrapolation

Extrapolative Ansätze zur Fertilitätsprognose schreiben den zu prognostizierenden Fertilitätsparameter in Abhängigkeit seiner historischen Entwicklung ohne Einschluss weiterer Informationen fort. Anstelle des Fertilitätsparameters können aber auch Modellparameter eines funktionalen Modells zur Bestimmung der Fertilitätsparameterwerte extrapoliert werden.

Grundsätzlich kann (wie bei der Mortalitätsprognose) zwischen der graphischen und der mathematischen Extrapolation unterschieden werden (vgl. Abschnitt 5.2.1 ab Seite 84). Die graphische Extrapolation ist aufgrund ihrer Subjektivität weder reproduzierbar noch reliabel, sodass sie im Vergleich zur mathematischen Extrapolation nachrangig einzuordnen ist.

### Konstante Fertilitätsannahmen

Eine einfache Möglichkeit der Fertilitätsprognose besteht darin, die Fertilität mit einem konstanten Wert in die Zukunft zu extrapolieren. Zur Bestimmung eines solchen konstanten Wertes kann beispielsweise der zuletzt beobachtete Wert oder auch ein Durchschnittswert zuletzt beobachteter Werte des jeweils zu prognostizierenden Fertilitätsparameters herangezogen wer-

---

<sup>6</sup>Umfassende Überblicke zu bestehenden Ansätzen zur Fertilitätsprognose geben beispielsweise de Beer [67], Smith et al. [256] und Booth [30].

den.<sup>7</sup> Die Annahme konstanter Werte für einen Fertilitätsparameter ist dann geeignet, wenn gleichbleibende Fertilitätsverhältnisse unterstellt werden können oder wenn die Prognose der Fertilität derart unsicher ist, dass weder ein Ansteigen noch ein Absinken mit mehr Sicherheit vorausgesagt werden kann. Darüber hinaus können auch Benchmark-Modelle konstante Fertilitätswerte verwenden, um z. B. eine Bevölkerungsentwicklung als Konsequenz bestimmter Annahmen zu prognostizieren. Soll jedoch der wahrscheinlichste Verlauf der zukünftigen Fertilität prognostiziert werden, ist die Annahme konstanter Fertilität nur solange geeignet bis die Annahme stabiler Fertilitätsbedingungen verletzt wird. Demzufolge eignet sich die Annahme konstanter Fertilität eher für kurze Projektionshorizonte [67, 256].

### Nichtkonstante Fertilitätsannahmen mit Trendextrapolationsmodellen

Nichtkonstante Fertilitätsannahmen können über Trendextrapolationsmodelle generiert werden, die einen funktionalen Zusammenhang zwischen dem zu prognostizierenden Fertilitätsparameter und der Zeit modellieren. Dabei hängt die zukünftige Entwicklung des Fertilitätsparameters vornehmlich von der Wahl des Trendextrapolationsmodells und der Länge der Basisperiode ab. So können mit der Wahl eines geeigneten Trendextrapolationsmodells subjektive Erwartungen des Prognostikers insofern in die Fertilitätsprognose einfließen, als die Stärke und die zeitliche Entwicklung der Anstiege<sup>8</sup> oder auch feste zu erreichende Niveaus<sup>9</sup> bestimmt werden können. Zudem werden mit den historischen Werten des zu prognostizierenden Fertilitätsparameters aus der frei wählbaren Basisperiode die Modellparameter des Trendextrapolationsmodells geschätzt, sodass der in der Basisperiode beobachtete Trend (ohne Einschluss weiterer Informationen) in die Zukunft fortgeschrieben wird. Dabei kann der fortzuschreibende Trend je nach auszuwählender Länge der Basisperiode variieren [67].

Häufig werden komplexe Zeitreihenmodelle nach Box und Jenkins [36] zur Prognose aggregierter und disaggregierter Fertilitätsparameter verwendet, da sie mit dem Einbau zufälliger Schwankungen Unsicherheitsbereiche (inform von Konfidenzintervallen) angeben können. So hat z. B. Saboia [243] bereits 1977 ein ARIMA(p,d,q)-Zeitreihenmodell (einer höheren Ordnung p) zur Prognose der TFR vorgeschlagen, das nicht nur kurz- sondern auch langzeitige

---

<sup>7</sup>So prognostiziert z. B. das Statistische Bundesamt in der 11. und 12. koordinierten Bevölkerungsprognose bis 2050 bzw. 2060 die TFR in einer von insgesamt drei verschiedenen Fertilitätsannahmen mit einem konstanten Wert von (annähernd) 1,4 [265, 267].

<sup>8</sup>z. B. mit der Wahl eines linearen oder exponentiellen Trendextrapolationsmodells

<sup>9</sup>z. B. mit der Wahl eines logistischen Trendextrapolationsmodells

Schwankungen, die die Länge des Generationenabstands repräsentieren, berücksichtigt. Nachfolgend sollen die extrapolativen Ansätze von Lee [155] und Hyndman und Ullah [118] genauer vorgestellt werden, da sie ebenfalls ein Zeitreihenmodell zur Prognose der Fertilität nutzen.

**Der Ansatz von Lee** Lee [155] hat 1993 eine Abwandlung des Lee-Carter Modells (vgl. Abschnitt 5.2.1 ab Seite 84) vorgeschlagen, um (logarithmierte) altersspezifische Fertilitätsraten,  $f_{x,t}$ , zu prognostizieren (vgl. Formel 6.1).

$$\ln f_{x,t} = \alpha_x + f_t \beta_x + \varepsilon_{x,t} \quad (6.1)$$

Der Modellparameter  $\alpha_x$  stellt dabei jeweils den Durchschnittswert der logarithmierten Fertilitätsraten eines Alters  $x$  in der Basisperiode dar.<sup>10</sup> Die Abweichungen der altersspezifischen Fertilitätsraten von ihrem jeweiligen Durchschnittswert  $\alpha_x$  werden von  $\beta_x$ <sup>11</sup> für jedes Alter und von dem Fertilitätsindex  $f_t$ <sup>12</sup> über die Zeit  $t$  ausgedrückt. Darüber hinaus stellen die  $\varepsilon_{x,t}$  die normalverteilten Abweichungen des Modells mit einem Mittelwert von 0 und einer konstanten Varianz dar, die die vom Modell nicht eingefangenen Abweichungen der modellierten von den tatsächlich beobachteten Werten wiedergeben.

Die Modellparameterwerte für  $\beta_x$  und  $f_t$  werden mit den zentrierten alters- und zeitspezifischen Fertilitätsraten<sup>13</sup> der Basisperiode über eine Singulärwertzerlegung geschätzt.

Anschließend wird der Fertilitätsindex,  $f_t$ , mit einem geeigneten Zeitreihenmodell, das an die mit der Singulärwertzerlegung geschätzten Werte des Fertilitätsindex angepasst wird, prognostiziert. Um möglichst plausible Werte für die zukünftige Fertilität zu prognostizieren, kann der Fertilitätsindex vereinfachend für die TFR (vgl. Formel 6.2) unter bestimmten (vom Prognostiker frei regulierbaren) Bedingungen extrapoliert werden.

$$\ln TFR = \ln \overline{TFR}_{Basisperiode} + f_t + \varepsilon_t \quad (6.2)$$

So können nach Lee [155] sowohl eine fixe obere und untere Grenze als auch ein ultimatives Niveau für die TFR über den gesamten Projektionshorizont vorgegeben werden. Bei der Fertilitätsprognose für die USA von 1990 bis 2065 soll beispielweise die TFR weder die 0

<sup>10</sup>Folglich stimmt die Summe aller  $\alpha_x$  (über alle Alter  $x$ ) mit der durchschnittlichen (logarithmierten) TFR der Basisperiode,  $\ln \overline{TFR}_{Basisperiode}$ , überein.

<sup>11</sup>Die Summe der  $\beta_x$  soll dabei über alle Alter  $x$  eins sein.

<sup>12</sup>Die Summe der  $f_t$  soll dabei über alle  $t$  der Basisperiode null sein.

<sup>13</sup>Eine zentrierte alters- und zeitspezifische Fertilitätsrate bestimmt sich aus der Differenz:  $f_{x,t} - \alpha_x$ .

unter- noch die 4 überschreiten und ein ultimatives Niveau von 1,85 anstreben.

Die prognostizierten Werte des Fertilitätsindex können dann in die Modellgleichung (vgl. Formel 6.1), eingesetzt werden, sodass im Zusammenschluss mit den Modellparameterwerten der  $a_x$  und  $\beta_x$  prognostizierte Werte für die (logarithmierten) altersspezifischen Fertilitätsraten resultieren.

Zwei der wesentlichen Vorteile dieses Ansatzes zur Fertilitätsprognose sind die relativ einfache Umsetzung wie auch die Berechnung von Konfidenzintervallen, wenn Zeitreihenmodelle zur Prognose des Fertilitätsindex verwendet werden. Nachteilig wirkt sich hingegen die Modellierung der inflexiblen Fertilitätsveränderung über die Alter und die Zeit aus. So ist beispielsweise die Modellierung einer erst stark und dann allmählich sinkenden Fertilität der jungen Frauen und eine erst leicht und dann stark zunehmende Fertilität der älteren Frauen mit diesem Ansatz nicht möglich, da die  $\beta_x$  zeit- und die  $f_t$  altersinvariant sind. Demzufolge eignet sich dieser Ansatz eher zur Kurz- als zur Langzeitprognose der Fertilität.<sup>14,15</sup>

**Der Ansatz von Hyndman und Ullah** Hyndman und Ullah [118] haben 2007 einen *functional data* Ansatz zur Prognose altersspezifischer Fertilitäts(- und Sterbe)raten vorgeschlagen, der auf dem Lee-Carter Modell [157, 50] zur Mortalitätsprognose und dem Ansatz von Lee [155] zur Fertilitätsprognose aufbaut (vgl. Abschnitt 5.2.1 ab Seite 88). Dabei werden die geglätteten und logarithmierten Fertilitätsraten,  $\ln f_{x,t}$ , derart flexibel prognostiziert, dass einzelne Altersbereiche, die durch Basisfunktionen,  $\phi_{k,x}$ , repräsentiert werden, separat über die  $\beta_{t,k}$  mit einem stochastischen Zeitreihenmodell fortgeschrieben werden können (vgl. Formel 5.5 auf Seite 89).<sup>16</sup> Dadurch wird die invariante Prognose der Fertilität über alle Alter und die Zeit von dem Modell von Lee [155] aufgehoben, sodass das Modell von Hyndman und Ullah [118] für längere Projektionshorizonte geeignet wäre, wenn nicht ausschließlich die historische Fertilitätsentwicklung als Informationsquelle genutzt werden würde. Denn so können keine neuartigen Fertilitätsentwicklungen prognostiziert werden, die mit anderen Informationsquellen (wie z. B. Expertenmeinung und Fertilitätstheorien (vgl. Abschnitt 6.1.3

---

<sup>14</sup>Dennoch nutzen Lee und Tuljapurkar [159] diesen Ansatz zur Generierung stochastischer Fertilitätsannahmen für eine probabilistische Bevölkerungsprognose (vgl. Abschnitt 4.2.3 ab Seite 63).

<sup>15</sup>Lee [152] beschreibt 2004 die Rückkehr zu einem stochastischen Ansatz zur Fertilitätsprognose als Niederlage all seiner vorangegangenen erfolglosen Versuche, andere, erklärende Methoden zur Fertilitätsprognose einzusetzen.

<sup>16</sup>Bei der Prognose der australischen Fertilität von 2001 bis 2020 ermitteln Hyndman und Ullah [118] drei Basisfunktionen, die die Fertilität in den hohen, niedrigen und mittleren reproduktiven Altern darstellen.

ab Seite 123)) antizipiert werden können.

Da Trendextrapolationsmodelle lediglich die vergangene Entwicklung in die Zukunft fort-schreiben, können sie keine Trendumkehrungen infolge instabiler fertilitätsbeeinflussender Faktoren prognostizieren. Folglich sind sie aufgrund ihrer Stabilitätsannahme nur für kurze Projektionshorizonte geeignet.

### **Fertilitätsgesetze**

Eine weitere Möglichkeit der extrapolativen Fertilitätsprognose besteht darin, die (beständi-ge) Verteilung der altersspezifischen Fertilitätsraten mit einem geeigneten Modell funktional nachzubilden. Dabei werden in einem ersten Schritt die Werte der Modellparameter (des jeweiligen Fertilitätsgesetzes) für jede Verteilung der altersspezifischen Fertilitätsraten aus der Basisperiode geschätzt. Anschließend können die zukünftigen Werte eines jeden Mo-dellparameters anhand seiner ermittelten Zeitreihenwerte aus der Basisperiode extrapoliert werden. Der Einsatz dieser zukünftigen Modellparameterwerte in das Fertilitätsgesetz ergibt dann die zukünftigen Werte der altersspezifischen Fertilitätsraten<sup>17</sup> [67, 170].

Es gibt eine Vielzahl von Modellen zur Beschreibung der glockenförmigen Verteilung der altersspezifischen Fertilitätsraten wie z.B. die Coale-Trussell Funktion [61], die Hadwiger Funktion, (un)modifizierte Beta- und Gammaverteilungen [133, 279], die Doppelexponen-tialfunktion von Rogers [237, 239], polynomiale Modelle oder auch Spline Funktionen (wie z.B. das quadratische Splinemodell von Schmertmann [248]).<sup>18</sup> Diese Modelle unterscheiden sich grundsätzlich in der Anzahl der zu berücksichtigenden Parameter wie auch in deren demographischer Interpretierbarkeit [56]. Hoem et al. [111] unterscheiden zudem zwei verschiedene Einsatzmöglichkeiten für Fertilitätsgesetze: Einerseits die Fertilitätsprognose und andererseits die Glättung tatsächlich beobachteter Verteilungen altersspezifischer Fertilitätsraten zur Erkennung und Analyse der jeweils dahinterstehenden, von zufälligen Abweichungen bereinigten Struktur. Fertilitätsgesetze, die sich zur Fertilitätsprognose eignen, sollten generell nur eine kleine Anzahl an Parametern verwenden, die obendrein sinnvoll demographisch interpretiert (und damit auch prognostiziert) werden können<sup>19</sup>

---

<sup>17</sup>Dieses Verfahren entspricht den Mortalitätsgesetzen zur Prognose der Sterblichkeit (vgl. Abschnitt 5.2.1 ab Seite 90).

<sup>18</sup>Einen umfassenden Überblick über bestehende Modelle zur Beschreibung der Verteilung altersspezifischer Fertilitätsraten geben z.B. Hoem et al. [111], Booth [30], Bell [18] und Peristera et al. [213].

<sup>19</sup>Dabei können die Parameter nicht nur über die Zeitreihe ihrer geschätzten Werte, sondern auch, bei entspre-

[140, 195, 237, 240, 279].

Hoem et al. [111] haben 1981 die Anpassungsfähigkeit diverser Fertilitäts*gesetze* für dänische Daten zur Fertilität in der Zeit von 1962 bis 1971 miteinander verglichen. Dabei zeigen eine Spline Funktion, die Hadwiger Funktion, eine modifizierte Gammaverteilung wie auch die Coale-Trussell Funktion eine relativ hohe Anpassung an die beobachteten Daten. Zu einer anderen Einschätzung führt hingegen die vergleichende Analyse der Anpassungsfähigkeit diverser Fertilitäts*gesetze* von Peristera et al. [213] für verschiedene Länder in dem Zeitraum von 1975 bis 2000. Hier erzielen die Betaverteilung und eine Spline Funktion deutlich genauere Anpassungen an tatsächlich beobachtete Fertilitätsverteilungen als die Hadwiger Funktion und die Gammaverteilung.

Demnach reagieren die Fertilitäts*gesetze* zwar relativ sensibel auf Veränderungen in der Fertilitätsverteilung. Allerdings sind sie auch derart inflexibel, dass sie die leicht von der nahezu symmetrischen Glockenkurve abweichenden Verteilungen der altersspezifischen Fertilitätsraten, wie sie in letzter Zeit in einigen entwickelten Ländern zu beobachten sind, nur ungenügend einfangen können. So zeigt sich beispielsweise in Irland, dem Vereinigten Königreich [56] oder auch in den USA [213] eine erhöhte Fertilität sowohl in den jungen als auch in den mittleren reproduktiven Altern der Frauen.<sup>20</sup> Um diese Entwicklungen angemessen einfangen zu können, schlagen Chandola et al. [56] eine Verbindung zweier Verteilungen, die jeweils der Hadwiger Funktion ähneln, über einen Parameter  $m$  vor, der das relative Gewicht der jungen und mittleren Fertilität zueinander bestimmt. Diese Vorgehensweise resultiert aus der Annahme, dass die erhöhte Fertilität in den jungen und mittleren reproduktiven Altern der nichtehelichen und ehelichen Fertilität in diesen Ländern entspräche. Daneben schlagen auch Peristera et al. [213] ein flexibles Modell mit 6 Parametern vor, um die neu auftretende Heterogenität in der Fertilität einiger Länder adäquat zu berücksichtigen. Die Parameter entsprechen dabei dem Niveau der Fertilität, dem mittleren Alter bei Geburt wie auch der Streuung in den jungen und mittleren reproduktiven Altern. Neben der ehelichen und nichtehelichen Fertilität führen Peristera et al. [213] noch weitere Faktoren wie die Bildung, den sozioökonomischen Status, die Religion oder auch

---

chender demographischer Bedeutung (wie z. B. der  $TFR$ , dem Durchschnittsalter oder auch dem Modalalter bei der Geburt eines Kindes) separat mit jeweils geeigneten Methoden (wie z. B. der Trendextrapolation) prognostiziert werden.

<sup>20</sup>Bei näherer Betrachtung der Abbildung 6.2 zeigt sich auch in den neuen Ländern im Jahr 2008 eine (lokale) Erhöhung in den jungen und mittleren reproduktiven Altersstufen.



die ethnische Zusammensetzung als Grund für die heterogene Fertilität auf. So vermuten sie, dass z.B. in Spanien eine hohe Zahl von Immigranten die Fertilitätsmuster aus den Herkunftsländern beibehält und entgegen der einheimischen Bevölkerung Geburten in relativ jungen Altern realisiert.<sup>21</sup> Ein Vergleich der flexiblen Modelle von Peristera et al. [213] und Chandola et al. [56] zeigt ähnlich hohe Anpassungswerte [213], sodass beide zur Modellierung der neu auftretenden Fertilitätsverteilungen geeignet sind.

Die Prognose der gesamten Verteilung der altersspezifischen Fertilitätsraten mit einem geeigneten *Fertilitätsgesetz* ist gegenüber der separaten, unabhängigen Prognose der einzelnen altersspezifischen Fertilitätsraten insofern vorteilhaft, dass eine glatte und vor allem konsistente Fertilitätsverteilung für die einzelnen Projektionsintervalle, insbesondere über längere Projektionshorizonte, resultiert [279]. Darüber hinaus wird der komplexe Prognoseprozess der einzelnen altersspezifischen Fertilitätsraten mit einem *Fertilitätsgesetz* auf die Prognose weniger, möglichst demographisch sinnvoll interpretierbarer Modellparameter reduziert. Allerdings können *Fertilitätsgesetze* meist nur die bereits in der Vergangenheit beobachteten Fertilitätsverteilungen erzeugen, sodass zukünftige, neuartige Veränderungen in der Verteilungsform (wie z.B. die Entwicklung einer zweiten lokalen Häufung der Fertilität in einem jüngeren Alter) nur ungenügend eingefangen werden können. Demzufolge eignen sich *Fertilitätsgesetze* eher zur Kurzzeitprognose der Fertilität. Für Langzeitprognosen eignen sich *Fertilitätsgesetze* folglich nur, wenn sie flexibel auf Veränderungen in der Verteilungsform reagieren.

### 6.1.2. Interpolation

Interpolative Ansätze zur Fertilitätsprognose schreiben den Entwicklungspfad (meist) zwischen dem zuletzt beobachteten und dem zu bestimmenden Zielwert eines zu prognostizierenden Fertilitätsparameters fort. Die Bestimmung des Zielwerts stellt dabei das wesentliche Kriterium zur Unterscheidung diverser interpolativer Ansätze zur Fertilitätsprognose dar.

Eine Möglichkeit zur Bestimmung eines ultimativen Fertilitätsniveaus besteht darin, die beobachteten Fertilitätsparameterwerte einer fortgeschritteneren (Sub)Population auf eine weniger fortgeschrittenere Bevölkerung mit einem gewissen Zeitabstand zu übertragen. Dies

---

<sup>21</sup>Eine nähere Beschreibung bestehender Theorien zur Erklärung des Fertilitätsverhaltens von Immigranten erfolgt im Abschnitt 9.1.6 ab Seite 172.

kann aufgrund der Annahme erfolgen, dass die fortgeschrittenere (Sub)Population den für die weniger fortgeschrittenere Bevölkerung (zum Zeitpunkt der Prognoseerstellung) wahrscheinlichsten Entwicklungsverlauf der Fertilität bereits erfahren hat, da für sie dieselben (bzw. hinreichend ähnliche) Fertilitätsbedingungen gegolten haben, die in der Zukunft für die weniger fortgeschrittenere Bevölkerung erwartet werden [67, 256].

Darüber hinaus können ultimative Fertilitätsniveaus mit Erhebungen zur erwarteten bzw. gewünschten Kinderzahl oder auch mit Expertenmeinung bestimmt werden [168, 170]. Lee [154] hat z.B. 1980 einen Ansatz zur Fertilitätsprognose vorgeschlagen, der das ultimative Fertilitätsniveau auf der Basis von zu erhebenden Kinderwunschzahlen bestimmt. Dabei kann das ultimative Fertilitätsniveau (inform einer gewünschten Kinderzahl) für jede Kohorte an die sich permanent verändernden sozioökonomischen Rahmenbedingungen angepasst werden. Lee [154] nimmt dazu vereinfachend an, dass die gewünschte Kinderzahl für alle Mitglieder einer Kohorte bzw. Altersgruppe identisch ist und synchron für alle Kohorten bzw. Altersgruppen über den Projektionshorizont (systematisch) schwankt. Außerdem wird der jährliche Wert des zu prognostizierenden Fertilitätsparameters als fixer Anteil der durchschnittlich noch zu erwartenden Kinderzahl dieser Altersgruppe bestimmt.<sup>22</sup> Mit diesen Modellannahmen kann dann der Verlauf des zu prognostizierenden Fertilitätsparameters flexibel (mit diversen Trendextrapolationsmodellen z.B. linear, logistisch oder auch (co)sinusförmig) an die zusätzlich (für jede Kohorte) zu prognostizierende *moving* Kinderwunschzahl angepasst werden. Ein Nachteil dieses relativ realistischen und flexiblen Ansatzes zur Fertilitätsprognose besteht darin, dass nicht nur der eigentliche Fertilitätsparameter, sondern auch noch die gewünschte Kinderzahl und der daran anzupassende (bzw. konvergierende) Verlauf des Fertilitätsparameters prognostiziert werden müssen. Da die gewünschte Kinderzahl von diversen (individuellen und äußeren) Faktoren abhängig ist, die von einem Prognostiker z.T. nur schwer vorauszusehen sind, ist dieser Ansatz mit einer erhöhten Prognoseunsicherheit verbunden.

Der Verlauf des zu prognostizierenden Fertilitätsparameters zwischen seinem Start- und Zielwert kann mit verschiedenen Methoden (wie z.B. einem Zeitreihenmodell oder auch Expertenmeinung) interpoliert werden.

---

<sup>22</sup>Für die USA ermittelt Lee [154] zur Bestimmung der jährlichen Geburtenrate in einem Alter  $x$  einen fixen Anteil von 0,18 in der Zeit von 1955 bis 1975.

Die interpolativen Ansätze können im Vergleich zu den extrapolativen Ansätzen zur Fertilitätsprognose neben der historischen Entwicklung zusätzliche Informationen (wie z. B. erklärende Theorien oder auch Ergebnisse aus entsprechenden Erhebungen) berücksichtigen, sodass das zukünftige Entwicklungsspektrum der Fertilität umfassender abgedeckt werden kann. Allerdings können sich aufgrund der z. T. unkonkreten Vorgaben (zu den einzelnen Arbeitsschritten) etliche Probleme bei der praktischen Umsetzung ergeben. So ist beispielsweise die Wahl einer geeigneten fortgeschritteneren (Sub)Population wie auch die Bestimmung der exakten zeitlichen Verzögerung zwischen den Fertilitätsniveaus der weniger fortgeschrittenen und der fortgeschritteneren (Sub)Population problematisch, da das Eintreten desselben Entwicklungspfad insbesondere aufgrund der zeitlichen Veränderung der fertilitätsbeeinflussenden Faktoren unsicher ist. Daneben ist der Einsatz eines solchen interpolativen Ansatzes auf weniger fortgeschrittenere Bevölkerungen beschränkt, da es für fortgeschrittenere Bevölkerungen keine (Sub)Population gibt, die den zukünftigen Entwicklungspfad der Fertilität bereits durchlaufen hat. Darüber hinaus kann mit Expertenmeinung zwar uneingeschränkt bzw. mit vielfältigen Informationen ein ultimatives Fertilitätsniveau bestimmt werden, allerdings stets zulasten der Reproduzierbarkeit. Desweiteren ist die Bestimmung eines ultimativen Fertilitätsniveaus in Abhängigkeit der zusätzlich zu prognostizierenden gewünschten bzw. erwarteten Kinderzahl mit zusätzlicher Prognoseunsicherheit verbunden, da durch die Berücksichtigung zusätzlicher, z. T. schwer einschätzbarer äußerer und individueller Faktoren die Komplexität des Prognoseprozesses gesteigert wird. Ungewiss ist dabei, ob dieser erhöhte Prognoseaufwand auch tatsächlich durch eine erhöhte Prognosegenauigkeit aufgewogen wird.

### 6.1.3. Erklärende theoretische bzw. Qualitative Ansätze

Erklärende Theorien zum zukünftigen Fertilitätsverhalten können entweder pur oder unterstützend als zusätzliche Informationsquelle bei einer Fertilitätsprognose eingesetzt werden.

Es gibt eine Vielzahl erklärender Theorien zum Fertilitätsverhalten, die sich in der Art und dem Umfang der zu berücksichtigenden fertilitätsbeeinflussenden Faktoren wie auch in der Untersuchungseinheit unterscheiden. So können z. B. ökonomische, soziale, kulturelle, psychologische, medizinische oder auch politische Einflussfaktoren bei der theoretischen Erklärung der zu realisierenden Fertilität auf der Mikro- oder Makroebene<sup>23</sup> berücksichtigt werden. Je nachdem, welche Erklärungsfaktoren bei einer Theorie berücksichtigt werden,

---

<sup>23</sup>Neben den Mikro- und Makromodellen gibt es z. B. auch so genannte Mehrebenenmodelle. Dieser Aspekt soll an dieser Stelle jedoch nicht vertieft werden.

können beispielsweise ökonomische von sozial-psychologischen Theorien unterschieden werden. Während bei ökonomischen Theorien vorrangig der Nutzen und die Kosten von Kindern betrachtet werden, beschäftigen sich sozial-psychologische Theorien mit den von den Kindern zu erfüllenden Erwartungen bzw. Wünschen der Eltern.<sup>24</sup>

Ein Beispiel für einen ökonomischen Erklärungsansatz stellt beispielsweise die *Easterlin Hypothese* dar. Easterlin [78] postuliert 1973 eine inverse Beziehung zwischen der Fertilität und der relativen Kohortengröße. Wenn demnach eine dem Bestand nach kleinere auf eine größere Kohorte folgt, bekomme diese infolge besserer ökonomischer Opportunitäten (z.B. durch erhöhte Arbeitsplatzangebote und höhere Einkommen) mehr Kinder. Daraus resultiere eine zyklische Fertilitätsentwicklung, da sich kleine und große Kohorten in der zeitlichen Abfolge abwechselten. Lee [160] hat 1976 Easterlins Hypothese mit empirischen Datenanalysen für die USA unterstützt, indem er sowohl für die CFR als auch für die TFR einen inversen Zusammenhang zur relativen Kohortengröße bzw. zur Relation der potentiell Erwerbstätigen und der Bevölkerungsgröße ermittelt hat. Infolgedessen befürwortet Lee [160] zwar die Prognose der Fertilität in Abhängigkeit zur Kohortengröße als Alternative zu bisherigen Methoden, hinterfragt allerdings auch u. a. die Beständigkeit dieses Zusammenhangs. Neben den ansonsten raren, eindeutigen empirischen Belegen [67, 241] werden häufig auch die ungenauen Angaben zur Fertilitätshöhe und dem Timing der Geburten kritisiert, da mit Easterlins Hypothese lediglich Vorhersagen zur Richtung der Abweichungen vom Langzeittrend getroffen werden [241, 256]. So geht Ryder [241] zwar auch von einem Zusammenhang zwischen den ökonomischen Rahmenbedingungen, dem reproduktiven Verhalten und der relativen Kohortengröße aus. Allerdings zieht er daraus zusätzlich Schlüsse über das Timing von Geburten. Je einfacher demnach der Eintritt ins Erwerbsleben sei, desto eher werden Geburten im Lebensverlauf realisiert und umgekehrt. Demgemäß werden in Abhängigkeit der relativen Kohortengrößen und der ökonomischen Bedingungen Geburten im Lebensverlauf auf- und vorgeschoben, sodass sich daraus jährliche Schwankungen für die TFR ergeben. Neben Annahmen zum Timing der Geburten trifft Ryder [241] auch Annahmen zum Langzeittrend der Fertilität (in Abhängigkeit diverser z.B. politischer, ökonomischer und normativer Faktoren), sodass daraus eine theoretisch fundierte Vorgehensweise für die Durchführung einer Fertilitätsprognose resultiert.

---

<sup>24</sup>Einen umfassenden Überblick über bestehende Theorien zur Erklärung reproduktiven Verhaltens geben z. B. de Bruijn [69, 70], van de Kaa [288], Schmitt [251] oder auch Ermisch [81].

Weitere, häufig in fertilitätserklärenden Theorien<sup>25</sup> angebrachte Faktoren sind beispielsweise die zunehmende Partizipation der Frauen am Bildungsprozess und auf dem Arbeitsmarkt, die Veränderung der Werte und Einstellungen, sodass sich (u. a. durch den zunehmend ausgelebten Willen zur Selbstverwirklichung) die Position der Frauen gegenüber den Männern gestärkt hat und neben die Ehe als dominierende traditionelle Lebensform (diverse) zusätzliche moderne Lebensformen wie beispielsweise die nichteheliche Lebensgemeinschaft getreten sind. Die Prognose der zukünftigen Entwicklung dieser Einflussfaktoren und deren Auswirkungen auf die Fertilität können unterstützend bei einer Fertilitätsprognose eingesetzt werden.

Erklärende theoretische Ansätze können zwar vielfältige fertilitätsbeeinflussende Faktoren bei einer Fertilitätsprognose berücksichtigen. Allerdings müssen dann nicht nur die Fertilitätsparameter, sondern auch deren, z. T. schwer einschätzbare Einflussfaktoren prognostiziert werden, wodurch die Komplexität und vor allem die Prognoseunsicherheit deutlich zunehmen.<sup>26</sup> Fertilitätstheorien können allerdings unterstützend bei der Wahl eines geeigneten Extrapolationsmodells oder/und bei der Bestimmung eines ultimativen Niveaus eingesetzt werden, indem sie das Ausmaß und die Richtung der zukünftigen Fertilität begründet vorgeben.

### 6.1.4. Methodenmix bzw. Multikausale Ansätze

Neben den extrapolativen, interpolativen und fertilitätstheoretischen Ansätzen gibt es auch multikausale Ansätze zur Fertilitätsprognose, die anstatt nur einer mehrere Informationsquellen und Methoden in sich vereinen. Ein wesentliches Ziel der multikausalen Ansätze zur Fertilitätsprognose besteht darin, die potentiellen Fehlerquellen angemessen einzuschätzen, sodass z. B. nach Long [168] die Genauigkeit der Prognoseergebnisse erhöht werden kann.

Alders et al. [3] beschreiben 2004 einen Ansatz zur Fertilitätsprognose von *Statistics Netherlands*, der auf der komplementierenden Analyse vergangener Fertilitätsprognosefehler, der Extrapolation des zu prognostizierenden Fertilitätsparameters mit einem geeigneten Zeitreihenmodell und Expertenmeinung basiert. Die erste Möglichkeit besteht darin, einfache Prognosefehler (vgl. Abschnitt 8.1.1 ab Seite 148) vergangener Fertilitätsprognosen auf den zu prognostizierenden Fertilitätsparameter zu übertragen oder z. B. bei einem Pro-

---

<sup>25</sup>wie z. B. die Theorie des zweiten demographischen Übergangs von van de Kaa [287]

<sup>26</sup>Dieses Problem ist bereits bei den Strukturmodellen im Abschnitt 3.2.4 ab Seite 47 beschrieben worden.

jektionshorizont, der die Länge der Projektionshorizonte vergangener Fertilitätsprognosen übertrifft, zu extrapolieren.<sup>27,28</sup> Eine weitere (modell-basierte) Möglichkeit besteht darin, die historische Zeitreihe eines zu prognostizierenden Fertilitätsparameters direkt mit einem Zeitreihenmodell, das automatisch Konfidenzintervalle generiert, zu extrapolieren. Bei beiden bisher beschriebenen Möglichkeiten zur Bestimmung eines Unsicherheitsbereichs für den zu prognostizierenden Fertilitätsparameter kann Expertenmeinung regulierend eingesetzt werden, indem bei der ersten Variante die zu berücksichtigenden vergangenen Fertilitätsprognosen und bei der zweiten Variante die Länge der Basisperiode subjektiv bestimmt werden. Alternativ kann Expertenmeinung in einer dritten Variante auch pur verwendet werden, sodass hauptsächlich auf der Basis paritätsspezifischer Fertilitätsanalysen, aber auch unter Berücksichtigung erhobener gewünschter bzw. erwarteter Kinderzahlen und weiterer fertilitätsbeeinflussender Faktoren (wie z. B. der Partizipation der Frauen am Arbeitsmarkt) ein Unsicherheitsbereich für den zu prognostizierenden Fertilitätsparameter bestimmt werden kann. Alle drei Varianten führen nach Alders et al. [3] zu einem ähnlichen 95-Prozent Konfidenzintervall für die Fertilität in den Niederlanden für das Zieljahr 2050, nach dem die TFR zwischen der unteren Grenze von 1,2 und der oberen Grenze von 2,3 mit einer Irrtumswahrscheinlichkeit von 5 Prozent liegen wird.

Ein multikausaler Ansatz zur Fertilitätsprognose wird z. B. auch in dem internationalen Projekt *UPE* (vgl. Abschnitt 4.2.5 ab Seite 73) genutzt, wobei hier neben der Analyse empirischer Prognosefehler, einem Zeitreihenmodell zur Extrapolation der Fertilität und Expertenmeinung zusätzlich Korrelationen der Prognosefehler zwischen den Ländern berücksichtigt werden [4, 7].

Da sich die verschiedenen bestehenden Ansätze zur Fertilitätsprognose für unterschiedlich lange Projektionshorizonte eignen, kann es sinnvoll sein, den Projektionshorizont einer Fertilitätsprognose mit einem multikausalen Ansatz so zu zerlegen, dass die zukünftige Fertilität in den jeweiligen Projektionsintervallen mit entsprechend geeigneten Methoden prognostiziert wird. So könnten bei den ersten Projektionsintervallen, die einem kurzen Projektionshorizont entsprechen, extrapolative Ansätze genutzt werden, wohingegen bei den

---

<sup>27</sup>Zur Extrapolation der Prognosefehler können Trendextrapolationsmodelle wie z. B. ein *Random Walk* verwendet werden.

<sup>28</sup>Ein  $(1 - \alpha)100$  Prozent Konfidenzintervall kann dabei so bestimmt werden, dass jeweils  $(\frac{\alpha}{2})100$  Prozent der aufsteigend sortierten Prognosefehler in einem Projektionsintervall über und unter der oberen und unteren Grenze liegen.

anschließenden Projektionsintervallen, die einem längeren Projektionshorizont entsprechen, Methoden eingesetzt werden könnten, die zunehmend zusätzliche Informationsquellen mit Expertenmeinung nutzen [67].

Multikausale Ansätze eignen sich in besonderem Maße zur Fertilitätsprognose, da sie versuchen, das zukünftige Entwicklungsspektrum der Fertilität umfassend mit weitestgehend allen zum Zeitpunkt der Prognoseerstellung verfügbaren Methoden und Informationsquellen abzudecken. Dadurch können sowohl die historische Entwicklung als auch Erwartungen zur zukünftigen Fertilitätsentwicklung, die mitunter auch einen Bruch mit der historischen Fertilitätsentwicklung darstellen, prognostiziert werden. So werden bei einem multikausalen Ansatz diverse potentielle Fertilitätsentwicklungen durch komplementierende Methoden und Informationsquellen eingefangen, wohingegen bei der Nutzung nur einer einzigen Methode bereits existierende Erkenntnisse unnötig bzw. vereinfachend vernachlässigt werden. Allerdings führt der Einsatz vielfältiger Methoden und Informationsquellen auch zu einer erhöhten Komplexität, die in einem ausgewogenen Maß zur (erhöhten) Genauigkeit der Fertilitätsprognose stehen sollte.

### 6.2. Schlussfolgerung — Vorschläge zur Generierung von Fertilitätsannahmen für eine Prognose mit dem PPPM

Dieses Kapitel zeigt die Vielzahl bestehender Ansätze zur Fertilitätsprognose, die sich hinsichtlich der Methode(n) und der zu berücksichtigenden Informationsquelle(n) unterscheiden. Extrapolative Ansätze schreiben die historische Entwicklung eines zu prognostizierenden Fertilitätsparameters fort, wohingegen interpolative, erklärende und multikausale Ansätze zusätzliche Informationsquellen (und damit auch von der vergangenen Fertilitätsentwicklung abweichende Verläufe) subjektiv mit einschließen.<sup>29</sup>

Je länger ein Projektionshorizont ist, desto wahrscheinlicher ist es, dass von der vergangenen Fertilitätsentwicklung abweichende Verläufe in der Zukunft (durch sich ständig verändernde fertilitätsbeeinflussende Faktoren) eintreten. Demnach können extrapolative Ansätze ausschließlich für kurze Projektionshorizonte angewendet werden, da dann die Stabilitätsannahme gleichbleibender Fertilitätsbedingungen in der Vergangenheit und

---

<sup>29</sup>Gleichwohl werden extrapolative Ansätze z. T. ergänzend auch in interpolativen, erklärenden und multikausalen Ansätzen genutzt.

Zukunft noch als (annähernd) realistisch einzuschätzen ist. Mit zunehmender Länge des Projektionshorizonts sollten jedoch Ansätze zur Fertilitätsprognose genutzt werden, die durch den subjektiven Einschluss zusätzlicher Informationsquellen (gegenüber der Vergangenheit) neuartige Fertilitätsentwicklungen modellieren und berücksichtigen können.

Neuartige Fertilitätsentwicklungen können demnach mit Ansätzen zur Fertilitätsprognose generiert werden, die Expertenmeinung zur Bestimmung eines ultimativen Niveaus oder auch zur Prognose der Auswirkung fertilitätsbeeinflussender Faktoren auf die zukünftige Fertilität nutzen. Umfassend kann das potentielle Spektrum der zukünftigen Fertilitätsentwicklung allerdings nur mit einem multikausalen Ansatz zur Fertilitätsprognose abgedeckt werden, indem zum Zeitpunkt der Prognoseerstellung weitestgehend alle verfügbaren Methoden und Informationsquellen sinnvoll eingesetzt werden. Damit sollte einerseits die Prognosegenauigkeit erhöht und andererseits die Prognoseunsicherheit reduziert werden.

Darüber hinaus sollte ein Fertilitätsparameter zur Prognose verwendet werden, der sensibel auf Veränderungen in der Altersverteilung der Fertilität reagiert und sowohl Perioden- und Kohorteneffekte als auch Veränderungen in der paritätsspezifischen Verteilung der Fertilität — wenn nicht direkt, dann doch zumindest indirekt — berücksichtigen kann. Im PPPM werden daher die altersspezifischen Fertilitätsraten als zu prognostizierender Fertilitätsparameter verwendet.

Bei der praktischen Anwendung des PPPM (vgl. Kapitel 12) werden zur Generierung der Annahmen für die altersspezifischen Fertilitätsraten vielfältige Informationsquellen mit geeigneten Methoden berücksichtigt.



## 7. Generierung von Migrationsannahmen

Die Migration ist von diversen politischen, sozialen und ökonomischen Faktoren der jeweiligen Ziel- und Herkunftsländer abhängig, sodass ihre Prognose (z. T. auch aufgrund ihrer vergleichsweise schwierigen Datenlage) besonders komplex und unsicher ist.

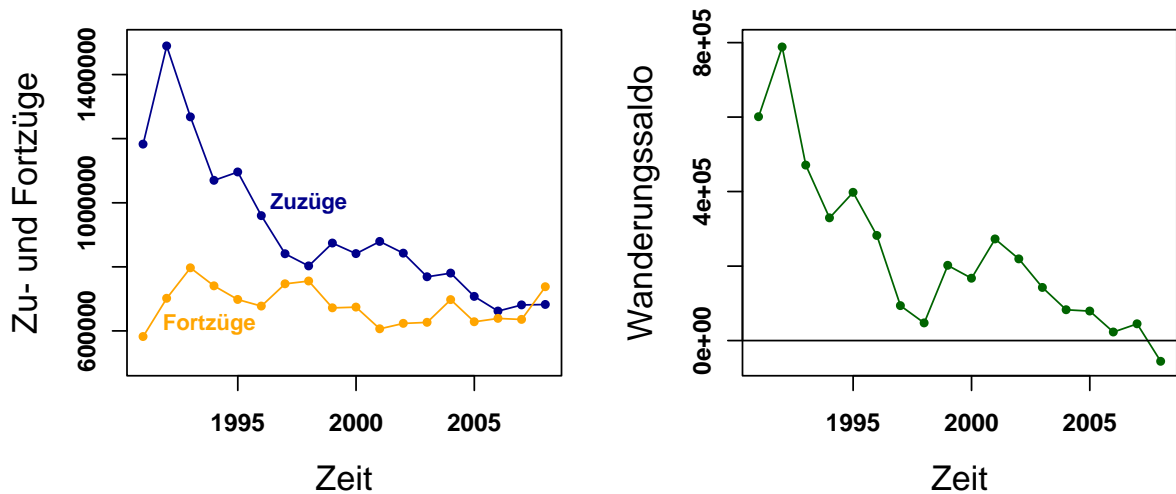


Abbildung 7.1.: Entwicklung der Zu- und Fortzüge (links) und des Wanderungssaldos (rechts) über die Außengrenzen Deutschlands in der Zeit von 1991 bis 2008; Datenquelle: Statistisches Bundesamt

Die Entwicklung der Zu- und Fortzüge (vgl. Abbildung 7.1) über die Außengrenzen Deutschlands ist relativ instabil.<sup>1</sup> So schwanken die Zuzüge im Zeitraum von 1991 bis 2008 zwischen 1.489.945 im Jahr 1992 und 661.855 im Jahr 2006 beträchtlich, wohingegen die Fortzüge in dieser Zeit um einen Wert von 679.997,8 schwanken. Die Instabilität der Wanderungen hängt von dem komplexen, sich permanent verändernden Zusammenspiel sozialer, ökonomischer und politisch(-rechtlich)er Bedingungen in den jeweiligen Herkunfts- und Zielländern ab, sodass temporär Personen aus verschiedenen Beweggründen wandern. In den 1990ern zählen beispielsweise (Spät)Aussiedler, Asylantragsteller, aber auch Flüchtlinge

<sup>1</sup>Die Altersstruktur der Zu- und Fortzüge zeigt die Abbildung 12.12 auf Seite 262 für das Jahr 2006.

(aus dem ehemaligen Jugoslawien) verstärkt zu den Zuzügen nach Deutschland [41]. Sofern sich nur ein Faktor (in nur einem Herkunfts- oder Zielland) in dem komplexen Zusammenspiel verändert, kann sich das Migrationsgeschehen global und lokal für die zu prognostizierende Bevölkerung entscheidend verändern.

Dennoch ist die Migration in der Vergangenheit oftmals bei Bevölkerungsprognosen ob ihrer vermeintlichen quantitativen Bedeutungslosigkeit oder/und aufgrund ihrer unvollständigen Datenreihen vernachlässigt worden<sup>2</sup> (vgl. Abschnitt 3.1.3 ab Seite 31). Allerdings ist die Prognose der Migration unerlässlich, da sie die einheimische Bevölkerung nicht nur vorübergehend quantitativ stärkt oder schwächt, sondern langfristig in ihrer Zusammensetzung nach Alter, Geschlecht, Ethnie und Kultur verändern kann. Folglich kann sich auch das Verhalten der zu prognostizierenden (Gesamt)Bevölkerung hinsichtlich wesentlicher demographischer Ereignisse (wie Geburten und Sterbefälle) infolge langfristiger Migration verändern. Dieser Aspekt tritt umso mehr in den Vordergrund, wenn die quantitative Bedeutung der natürlichen Bevölkerungsbewegung für das Bevölkerungswachstum der einheimischen Bevölkerung (z. B. infolge eines drastischen Fertilitätsrückgangs) abnimmt [21, 68, 94, 114, 289].

Demzufolge sollte ein geeigneter Parameter zur Prognose der Migration die Alters- und Geschlechtsstruktur einzelner Migranten-Subpopulationen (wie die direkten Zu- und Fortzüge und deren einzelne Nachkommengenerationen), wenn nicht sogar noch deren Herkunft, berücksichtigen, sodass die bevölkerungsdynamischen Auswirkungen von Migration auf die einheimische Bevölkerung detailliert prognostiziert werden können. Crujisen und Keilman [66] schlagen vor, die Migrationsereignisse (wie bei der Prognose der Fertilität und Mortalität) auf die jeweilige Risikobevölkerung zu beziehen. Allerdings entspricht die Risikobevölkerung der Immigranten im Prinzip der Weltbevölkerung, sodass die Berechnung und Prognose von Immigrationsraten zu aufwendig und mit den verfügbaren Daten unmöglich ist. Aus diesem Grund schlagen sie die Verwendung totaler Immigrationszahlen vor. Analog eignen sich zur Prognose der Emigranten totale Emigrationszahlen oder auch nach Crujisen und Keilman [66] Emigrationsraten, bei denen die Emigranten auf die jeweilige Risikobevölkerung, also auf die zu prognostizierende Bevölkerung bezogen wird. Im Gegensatz dazu ist die häufig in Bevölkerungsprognosemodellen verwendete Nettomigration (vgl. Abschnitt 4.2

---

<sup>2</sup>So hat z. B. das Statistische Bundesamt bei der ersten [252, 262] und zweiten [263, 318] koordinierten Bevölkerungsprognose keine Annahmen zur Migration getroffen [119, 271].

ab Seite 57) als zu prognostizierender Migrationsparameter ungeeignet, da sie aufgrund der vernachlässigten Unterschiede bei den Zu- und Fortzügen z. B. hinsichtlich Anzahl, Struktur (Alter, Geschlecht) und demographischem Verhalten (Fertilität und Mortalität) zu verzerrten Prognossergebnissen führen kann [22, 68].

## 7.1. Überblick über verschiedene Ansätze zur Migrationsprognose

Es gibt eine Vielzahl unterschiedlicher Ansätze zur Prognose der Migration, die sich in dem Aggregationsniveau des Prognosegegenstands wie auch in der Art und dem Umfang der zu nutzenden Informationsquellen unterscheiden. So sind extrapolative, von interpolativen, (theoretisch) erklärenden, bayesianischen und weiteren multikausalen Ansätzen zu unterscheiden, die entweder die Brutto- oder die Nettomigration prognostizieren.<sup>3</sup>

Da es für etliche Länder schwierig ist, detaillierte Migrationsdaten zu erhalten, behandeln die meisten Prognoseansätze die Migration relativ oberflächlich, indem nur die Nettomigration oder in Ausnahmefällen die Bruttomigration nach Alter und Geschlecht prognostiziert wird (vgl. Abschnitt 4.2 ab Seite 57). Allerdings sollen im (in dieser Arbeit zu entwickelnden) PPPM neben den direkten Zu- und Fortzügen auch deren einzelne Nachkommengenerationen separat prognostiziert werden, um u. a. detaillierte bevölkerungsdynamische Auswirkungen der Migration auf die einheimische Bevölkerung analysieren zu können (vgl. Abschnitte 9.1 und 12.4 ab den Seiten 168 und 242). Deshalb werden in diesem Kapitel zusätzlich Ansätze zur Fertilitäts- und Mortalitätsprognose für die einzelnen Subpopulationen der Migranten vorgestellt.

### 7.1.1. Ansätze zur Prognose der Bruttomigration

In diesem Abschnitt soll die Vielfalt der bestehenden Ansätze zur Prognose der Zu- und Fortzüge über die Außengrenzen eines Landes mit einigen ausgewählten Ansätzen vorgestellt werden.

---

<sup>3</sup>Einen umfassenden Überblick über bestehende Ansätze zur Migrationsprognose geben beispielsweise Raymer und Willekens [227], Bijak [21], Smith et al. [256], George et al. [94] und Booth [30].

### Extrapolative Ansätze

Extrapolative Ansätze schreiben zur Prognose der Migration die in der Vergangenheit beobachtete Migrationsentwicklung in die Zukunft fort.

Eine einfache Möglichkeit der extrapolativen Migrationsprognose besteht darin, den zuletzt beobachteten Wert oder auch einen Durchschnittswert vergangener Jahre konstant für den zu prognostizierenden Migrationsparameter in der Zukunft anzunehmen. Häufig wird auch vereinfachend die Nettomigration mit einem Wert von *null* fortgeschrieben<sup>4</sup> [68, 94, 289].

Eine weitere Möglichkeit besteht darin, die in der Basisperiode beobachtete Entwicklung des zu prognostizierenden Migrationsparameters mit einem geeigneten Trendextrapolationsmodell (wie z. B. mit einem ARIMA-Zeitreihenmodell) fortzuschreiben. Die Prognoseergebnisse können dabei durch die Wahl des Trendextrapolationsmodells und der Basisperiode subjektiv vom Prognostiker beeinflusst werden [22, 68].

Extrapolative Ansätze eignen sich bei der Prognose der Migration — wenn überhaupt — nur für kurze Projektionshorizonte, da die Annahme stabiler sozialer, ökonomischer und politisch(-rechtlich)er Bedingungen in den jeweiligen Herkunfts- und Zielländern über einen längeren Zeitraum unrealistisch ist und vergangene Migrationsentwicklungen auch kurzfristig (aufgrund *neuartiger* Bedingungen) große Schwankungen aufweisen können. Darüber hinaus ist der Einsatz einer so simplen Prognosemethode für die Migration nur insofern zu rechtfertigen, als die Berücksichtigung aller Einflussfaktoren (und deren zukünftige Entwicklung) unmöglich ist und folglich die Frage aufgeworfen werden kann, ob der Einschluss mehrerer verfügbarer Informationen überhaupt eine genauere Migrationsprognose erzeugt. Allerdings sollten die verfügbaren Informationen grundsätzlich genutzt werden, um die vielfältigen potentiellen Entwicklungsverläufe der Migration überhaupt aufdecken zu können und begründet — möglicherweise mit verschiedenen Annahmenverläufen — in einer Bevölkerungsprognose zu berücksichtigen.

---

<sup>4</sup>Eine Nettomigration von *null* kann entweder durch eine bewusste Vernachlässigung der Migration oder als zufälliges Resultat zu treffender Annahmen zu den Zu- und Fortzügen entstehen. Bei der zweiten Variante können daher dennoch (im Gegensatz zur ersten Variante) realistische bevölkerungsdynamische Effekte durch Veränderungen in der Alters- und Geschlechtsstruktur der zu prognostizierenden Bevölkerung durch Migration hervorgerufen werden.

### Interpolative Ansätze

Interpolative Ansätze schreiben die Entwicklung des zu prognostizierenden Migrationsparameters zwischen einem Startwert und einem mit vielfältigen Informationen zu bestimmenden Zielwert fort.

Der Zielwert des zu prognostizierenden Migrationsparameters kann dabei mit Expertenmeinung unter Berücksichtigung diverser Informationsquellen wie z.B. zu erwartender (inter)nationaler Migrationspolitiken und der vergangenen Migrationsentwicklung bestimmt werden [94]. Im Anschluss daran kann der konkrete Entwicklungspfad zwischen dem Start- und Zielwert des zu prognostizierenden Migrationsparameters entweder mit einem mathematischen Modell oder graphisch ermittelt (bzw. interpoliert) werden.

Interpolative Ansätze eignen sich nicht nur für kurz-, sondern auch für mittel- bis langfristige Migrationsprognosen, da sie zusätzliche Informationen berücksichtigen. Allerdings gibt es (wie bei den interpolativen Ansätzen zur Prognose der Mortalität und Fertilität (vgl. Abschnitte 5.2.2 und 6.1.2 ab den Seiten 94 und 121)) keine festen Vorgaben zur Bestimmung der Werte zwischen dem Start- und Zielwert, sodass diese entsprechend den Erwartungen eines Prognostikers variabel mit geeigneten Methoden modelliert werden können.

### Erklärende theoretische bzw. Qualitative Ansätze

Erklärende theoretische Ansätze können neben der vergangenen Migrationsentwicklung noch weitere, vielfältige Erklärungsfaktoren zur Migrationsprognose einsetzen.

Es gibt eine Vielzahl bestehender Theorien zur Erklärung internationaler (und nationaler) Migrationsbewegungen, die (unterstützend) zur Migrationsprognose verwendet werden können. Bijak [21, 23] unterscheidet soziologische, ökonomische, geographische und interdisziplinäre Migrationstheorien.<sup>5</sup> Prinzipiell versuchen diese Theorien wesentliche Erklärungsfaktoren für das Entstehen von Wanderung auf der Mikro- oder/und Makroebene herauszufiltern. Dabei wird grundsätzlich von der Annahme ausgegangen, dass Personen wandern, wenn sie erwarten, dadurch ihre Position (in einem oder mehreren Lebensbereichen)<sup>6</sup> verbessern zu können. Die Präferenzstruktur der zu verändernden Lebensbereiche wie

---

<sup>5</sup>Einen Überblick über bestehende Migrationstheorien geben z.B. auch Düvell [77], Howe und Jackson [114], Bähr [45] und Massey [184].

<sup>6</sup>— vorerst unabhängig davon, ob sie nun freiwillig oder unfreiwillig wandern —

auch die Art und der Umfang der für diesen Entscheidungsprozess zu berücksichtigenden Faktoren stellen wesentliche Unterscheidungsmerkmale der verschiedenen Migrationstheorien dar. So legen ökonomische Theorien z.B. den Fokus auf die ökonomische Position bzw. Situation einer (potentiell wandernden) Person bzw. ihres Haushalts (oder ihrer Familie), die versucht, ihr Einkommen zu maximieren bzw. (auch in der Zukunft) zu sichern. Neben dem Lohnniveau und dem Arbeitsplatzangebot können deshalb z.B. auch der ökonomische Entwicklungsstand wie auch der Diversifikationsgrad des Arbeitsmarktes in den jeweiligen Herkunfts- und potentiellen Zielbevölkerungen als Entscheidungsfaktoren herangezogen werden. Soziologische Migrationstheorien können im Vergleich dazu zwischen freiwilligen und unfreiwilligen Migrationsbewegungen unterscheiden. Die Verbesserung der ökonomischen Situation im Sinne einer Einkommensmaximierung ist z.B. als freiwillige Migrationsbewegung einzustufen, wohingegen die durch Missstände wie z.B. Kriege, Umweltkatastrophen, politische oder religiöse Verfolgung, Vertreibung und Armut ausgelöste Wanderung als unfreiwillig bzw. erzwungen einzustufen ist.<sup>7</sup> Darüber hinaus können soziologische Migrationstheorien auch die Präsenz dicht ausgebauter sozialer Netzwerke, Institutionen und Organisationen als wanderungsunterstützende Elemente bei der Erklärung von Migrationsbewegungen nutzen. Geographische Migrationstheorien berücksichtigen dagegen u.a. die Bevölkerungsgröße (z.B. im Zusammenhang mit dem Arbeitsangebot und dem Lohnniveau) und die räumliche Distanz (z.B. im Zusammenhang mit der Erreichbarkeit mittels diverser Transportmittel) zwischen der Herkunfts- und Zielregion als Erklärungsfaktoren für internationale Wanderungsbewegungen.<sup>8</sup> Im Vergleich zu den ökonomischen, soziologischen und geographischen Migrationstheorien versuchen interdisziplinäre Migrationstheorien die internationale Migration als Zusammenspiel diverser soziologischer, ökonomischer, geographischer und mitunter auch weiterer (wie beispielsweise psychologischer) Faktoren umfassend zu erklären.

Migrationstheorien können (pur) über Expertenmeinung oder auch (unterstützend) in mathematischen Modellen (wie z.B. Regressionsmodellen) zur Migrationsprognose genutzt werden. Da Migrationstheorien meist nur einzelne Aspekte des Entstehens internationaler Migration betrachten, können sie zur Steigerung der Prognosegenauigkeit (in Abhängigkeit der jeweils dominierenden Wanderungsgründe) komplementär für unterschiedliche Projektionsintervalle oder/und Subpopulationen eingesetzt werden [68].

---

<sup>7</sup>Der Ansatz der Push- und Pull-Faktoren geht auf Lee [150] aus dem Jahr 1966 zurück.

<sup>8</sup>Ravenstein [225, 226] hat bereits zum Ende des 19. Jahrhunderts bei der Ausarbeitung der ersten *Migrationsgesetze* die Bevölkerungsgröße und die räumliche Distanz zwischen der Herkunfts- und Zielregion als Erklärungsfaktoren für Migrationsbewegungen verwendet.

Allerdings sollten nicht nur verschiedene Migrationstheorien ergänzend, sondern auch noch zusätzliche, bislang (in den Theorien) vernachlässigte Aspekte wie z. B. die Steuerung internationaler Zu- und Fortzüge durch länderspezifische Migrationspolitiken oder auch die Entwicklung politischer (In)Stabilitäten über Expertenmeinung zusätzlich berücksichtigt werden [23, 149]. Der Einschluss weiterer Erklärungsfaktoren erhöht jedoch die Komplexität der Migrationstheorien derart, dass deren Einsatz als unterstützende Informationsquelle in einer Migrationsprognose immer schwieriger umzusetzen ist; zumal die theoretisch fundierten Zusammenhänge empirisch gemessen und ebenfalls mit geeigneten Methoden prognostiziert werden müssen. Dennoch eignen sich theoretisch erklärende Ansätze für langfristige Migrationsprognosen, da sie neuartige Entwicklungen unter Berücksichtigung vielfältiger Einflussfaktoren differenziert (z. B. nach Projektionsintervall und Subpopulation) antizipieren und prognostizieren können.

### **Multikausale Ansätze**

Für eine umfassende und genaue Prognose der (aus diversen Gründen entstehenden) Migration sollten vielfältige Informationsquellen sinnvoll kombiniert und in einem entsprechenden Modell angemessen operationalisiert werden können.

**Der multiregionale Ansatz von Rogers** Rogers [219, 234, 235, 236] hat ab den 1960ern ein multi- bzw. interregionales Modell zur Prognose einer Bevölkerung vorgeschlagen (und fortlaufend weiterentwickelt), das auf der *multistate* Kohorten-Komponenten Methode (vgl. Abschnitt 3.2.1 auf Seite 40) basiert und mit einer entsprechend erweiterten Projektionsmatrix (vgl. 3.2.1 ab Seite 35) umgesetzt werden kann. Dieses interregionale Modell kann neben dem Alter (und dem Geschlecht) auch noch die Lokalität bei einer Bevölkerungsprognose berücksichtigen, sodass die Migration nicht extern zur zu prognostizierenden Bevölkerung addiert (vgl. Formeln 3.2 bis 3.4 ab Seite 38), sondern integrativ während der Prognose inform kombinierter Migrations- und Mortalitätsraten berücksichtigt wird.

Um eine Bevölkerung (vereinfachend nur) nach dem Alter und der Region zu prognostizieren, müssen die Annahmen für die Fertilität, Mortalität und Migration alters- und regionalspezifisch sein. Werden allein zwei verschiedene Regionen A und B unterschieden, werden in dem anfänglichen Modell von Rogers [234] Annahmen für altersspezifische Fertilitätsraten pro Region ( $f_{x,A}$ ,  $f_{x,B}$ ), altersspezifische Überlebenswahrscheinlichkeiten pro

Region  $(p_{x,A}, p_{x,B})$  wie auch kombinierte altersspezifische Emigrations- und Mortalitätsraten aus der einen in die andere Region  $(e_{x,AB}, e_{x,BA})$ <sup>9</sup> berücksichtigt (vgl. Abbildung 7.2).

$$\begin{pmatrix} P_{x,t+1,A} \\ P_{x+1,t+1,A} \\ P_{x+2,t+1,A} \\ \vdots \\ P_{x,t+1,B} \\ P_{x+1,t+1,B} \\ P_{x+2,t+1,B} \\ \vdots \end{pmatrix} = \begin{pmatrix} f_{x,A} & f_{x+1,A} & \dots & 0 & 0 & 0 \\ {}_1p_{x,A} & 0 & 0 & {}_1e_{x,BA} & 0 & 0 \\ 0 & {}_1p_{x+1,A} & 0 & 0 & {}_1e_{x+1,BA} & 0 \\ 0 & 0 & \ddots & 0 & 0 & \ddots \\ 0 & 0 & 0 & f_{x,B} & f_{x+1,B} & \dots \\ {}_1e_{x,AB} & 0 & 0 & {}_1p_{x,B} & 0 & 0 \\ 0 & {}_1e_{x+1,AB} & 0 & 0 & {}_1p_{x+1,B} & 0 \\ 0 & 0 & \ddots & 0 & 0 & \ddots \end{pmatrix} \cdot \begin{pmatrix} P_{x,t,A} \\ P_{x+1,t,A} \\ P_{x+2,t,A} \\ \vdots \\ P_{x,t,B} \\ P_{x+1,t,B} \\ P_{x+2,t,B} \\ \vdots \end{pmatrix}$$

Abbildung 7.2.: Matrixnotation eines interregionalen Bevölkerungsprognosemodells nach Rogers [234] mit zwei Regionen A und B

Daneben hat Rogers [219, 235, 236] dieses interregionale Modell auch noch so formuliert, dass zusätzlich explizit Annahmen zur Fertilität und kombinierter Mortalität und Migration für Personen getroffen werden können, die aus der einen in die andere Region wandern (vgl. Abbildung 7.3).<sup>10</sup>

Die entsprechenden alters- und regionalspezifischen Fertilitäts- und Mortalitäts-/Migrationsraten können mit diversen Methoden wie z.B. der Extrapolation, der Interpolation oder auch Expertenmeinung prognostiziert werden.

Im Vergleich zu herkömmlichen Modellen (vgl. Abschnitt 4.2 ab Seite 57) berücksichtigt das interregionale Bevölkerungsprognosemodell von Rogers die Migration sehr detailliert. Zwar können pro Region separate Annahmen zur Fertilität, Mortalität und Migration für die Einheimischen und die im jeweils aktuellen Projektionsintervall auftretenden Emigranten getroffen werden, allerdings gelten dann (jeweils ab dem nächsten Projektionsintervall) diesel-

<sup>9</sup>Die  $e_{x,AB}$  entsprechen dabei dem Anteil der Personen, die aus der Region A in die Region B wandern und vom Alter  $x$  zum Alter  $x + 1$  überleben.

<sup>10</sup>So werden mit den  $f_{15,AA}$  die sich zum Zeitpunkt  $t + 1$  in der Region A aufhaltenden Geburten 15-jähriger Frauen, die sich auch zum Zeitpunkt  $t$  in der Region A aufgehalten haben, berechnet. Analog dazu werden mit den  $f_{15,BA}$  die sich zum Zeitpunkt  $t + 1$  in der Region A aufhaltenden Geburten 15-jähriger Frauen, die sich allerdings zum Zeitpunkt  $t$  noch in der Region B aufgehalten haben, berechnet.



$$\begin{pmatrix} P_{0,t+1,A} \\ P_{0,t+1,B} \\ \vdots \\ P_{14,t+1,A} \\ P_{14,t+1,B} \\ P_{15,t+1,A} \\ P_{15,t+1,B} \\ \vdots \\ P_{100+,t+1,A} \\ P_{100+,t+1,B} \end{pmatrix} = \begin{pmatrix} 0 & 0 & \dots & f_{14,AA} & f_{14,BA} & f_{15,AA} & f_{15,BA} & \dots & 0 & 0 \\ 0 & 0 & \dots & f_{14,AB} & f_{14,BB} & f_{15,AB} & f_{15,BB} & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 1P_{14,AA} & 1P_{14,BA} & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 1P_{14,AB} & 1P_{14,BB} & 0 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & 0 & 0 & 0 & \dots & 1P_{100+,AA} & 1P_{100+,BA} \\ 0 & 0 & \dots & 0 & 0 & 0 & 0 & \dots & 1P_{100+,AB} & 1P_{100+,BB} \end{pmatrix} \cdot \begin{pmatrix} P_{0,t,A} \\ P_{0,t,B} \\ \vdots \\ P_{14,t,A} \\ P_{14,t,B} \\ P_{15,t,A} \\ P_{15,t,B} \\ \vdots \\ P_{100+,t,A} \\ P_{100+,t,B} \end{pmatrix}$$

Abbildung 7.3.: Matrixnotation eines interregionalen Bevölkerungsprognosemodells nach Rogers [219, 235, 236] mit zwei Regionen A und B

ben Annahmen für die Einheimischen und die Immigranten in jeder Region. Folglich können in diesem Modell nicht die Einheimischen, Migranten und deren einzelne Nachkommengenerationen unterschieden werden, sodass z. B. auch nicht sukzessive Anpassungsprozesse in der Fertilität und Mortalität zwischen den Migranten und Einheimischen über die einzelnen Nachkommengenerationen modelliert werden können (vgl. Abschnitt 7.1.2 ab Seite 138). Da die Annahmen zur Fertilität, Mortalität und Migration mit diversen Methoden generiert werden können, gibt es diesbezüglich auch keine Einschränkungen hinsichtlich der zu nutzenden Informationsquellen. Folglich können mit dem interregionalen Modell von Rogers auch längerfristige Prognosen erstellt werden, wobei die Prognoseergebnisse umso mehr verzerrt werden, desto stärker sich das demographische Verhalten zwischen Migranten und Einheimischen auch nach dem Wanderungszeitpunkt unterscheidet.

**Der Ansatz von Bijak** 2008 schlägt Bijak [22] einen bayesianischen Ansatz zur Prognose internationaler Migration vor, bei dem die beiden Methoden Expertenmeinung und Zeitreihenanalyse formal miteinander kombiniert werden. Generell wird dabei die Wahrscheinlichkeitsverteilung (*posterior*) des zu prognostizierenden Migrationsparameters über ein (mathematisches) Prognosemodell, das sowohl auf den bereits beobachteten Daten der zu berücksichtigenden Modellparameter (*likelihood*) als auch auf (der daran sukzessive anzupassenden) Expertenmeinung (*prior*) basiert, bestimmt (vgl. Formel 5.10 auf Seite 108). Bijak [22] vergleicht zur (bayesianischen) Prognose der (logarithmierten) Emigrationsraten von Deutschland nach Polen und von Polen nach Deutschland für den kurzen Projektionshorizont von 2004 bis 2010 drei komplexe Zeitreihenmodelle, die sich in der Berücksichtigung ökonomischer Einflussfaktoren wie z. B. der Einkommensdifferenz und der Arbeitslosenquote

und den entsprechenden subjektiven Einschätzungen zur Ausprägung und Entwicklung der jeweiligen Parameterwerte unterscheiden. 100.000 Prognosedurchläufe erzeugen für jedes Projektionsintervall eine Wahrscheinlichkeitsverteilung für die zu prognostizierenden Emigrationsraten. Der Vergleich der drei Prognosemodelle zeigt, dass der Einsatz zusätzlicher Erklärungsfaktoren in Verbindung mit Expertenmeinung die Prognoseunsicherheit reduziert, indem die Breite der Wahrscheinlichkeitsverteilung für die zu prognostizierenden Emigrationsraten (zumindest in diesem Beispiel) abnimmt.

Multikausale Ansätze eignen sich besonders zur Prognose der instabilen Migration, da sie heterogene potentielle zukünftige Entwicklungen der Migration mit unterschiedlichen Methoden und Informationsquellen (auch über einen längeren Projektionshorizont) aufdecken können. Dabei sollten heterogene Migrationsentwicklungen mit verschiedenen Annahmenverläufen modelliert und in einer Bevölkerungsprognose berücksichtigt werden können.

### **7.1.2. Ansätze zur Prognose der Fertilität und Mortalität einzelner Migranten-Subpopulationen**

Im Gegensatz zu bestehenden Ansätzen zur Bevölkerungsprognose sollen im (in dieser Arbeit zu entwickelnden) PPPM die Immigranten und die Emigranten derart detailliert modelliert und prognostiziert werden, dass sowohl für die direkten Zu- und Fortzüge als auch für deren einzelne Nachkommengenerationen separat Annahmen zur Fertilität und Mortalität getroffen werden können (vgl. Abschnitt 9.1 und 12.4 ab den Seiten 168 und 242). Erst diese separate Betrachtung einzelner Subpopulationen erlaubt eine flexible Nachbildung sukzessiver Anpassungsprozesse der Immigranten und Emigranten an die Gegebenheiten im jeweiligen Zielland (oder auch das Beibehalten des demographischen Verhaltens aus dem jeweiligen Herkunftsland), sodass auch deren (bisher meist in Bevölkerungsprognosen ausgeblendete) bevölkerungsdynamische Konsequenzen umfassend berücksichtigt und analysiert werden können [257].

#### **Fertilitätsprognose**

Es gibt verschiedene Hypothesen, die zur Erklärung des fertilen Verhaltens von Migranten herangezogen werden können [107, 193, 257]. Die Hypothese der *Sozialisation* geht davon aus, dass die Fertilität der Migranten vorrangig von den in der Kindheit (bzw. im jeweiligen Herkunftsland) vermittelten Normen und Werten determiniert wird. Im Gegensatz dazu geht die Hypothese der *Adaption* davon aus, dass sich das Fertilitätsverhalten der Migranten

(mit zunehmender Aufenthaltsdauer) dem der Bevölkerung im Zielland anpasst, sodass das gegenwärtige Umfeld maßgeblich das Fertilitätsverhalten der Migranten bestimmt. Während das Fertilitätsverhalten der Migranten mit den Hypothesen der Sozialisation und Adaption langfristig erklärt werden soll, fokussieren sich die Hypothesen der Disruption und der Familien- bzw. Haushaltsgründung auf die kurzfristigen Auswirkungen der Migration auf die Fertilität. So geht die Hypothese der *Disruption* davon aus, dass die Fertilität während und kurz nach dem Wanderungsprozess durch den damit verbundenen Stress und die Umgewöhnung an neue soziale, ökonomische Bedingungen kurz unterbrochen wird, um sich anschließend an die Fertilität der Einheimischen anzupassen. Hingegen geht die Hypothese der *Familien- bzw. Haushaltsgründung* (mit einhergehender Eheschließung) von einer kurzzeitigen Fertilitätserhöhung unmittelbar nach dem Wanderungszeitpunkt aus, da bei Personen, die aufgrund einer Familienzusammenführung wandern, diese fertilitätsbegünstigenden Ereignisse zusammenfallen. Daneben geht die Hypothese der *Selektion* davon aus, dass Migranten eine selektive Gruppe hinsichtlich sozialer, ökonomischer und demographischer Merkmale darstellen, deren Fertilität von der Bevölkerung im Herkunftsland wie auch von der im Zielland abweichen kann. Sofern sich diese Merkmale zwischen Migranten und Einheimischen anpassen, können auch die Unterschiede in der Fertilität abnehmen.

Diverse empirische Untersuchungen zur Fertilität von Migranten im Zielland (respektive in Europa, den USA, Kanada, Brasilien und Australien) haben gezeigt, dass sich die Migranten mit zunehmender Aufenthaltsdauer wie auch über die Nachkommengenerationen dem Fertilitätsverhalten der einheimischen Bevölkerung im Zielland sukzessive anpassen [1, 17, 48, 74, 79, 107, 122, 123, 192, 193, 257, 320]. Milewski [192, 193] hat darüber hinaus festgestellt, dass sich die Fertilität der Migranten in Deutschland kurzfristig (im Sinne der Hypothese der Familien- bzw. Haushaltsgründung) erhöht und langfristig (im Sinne der Hypothese der Adaption) an die Fertilität der Einheimischen anpasst.

Allerdings stellen die Immigranten keine homogene Subpopulation dar. So können z. B. die Ethnie [120, 122, 193, 320], sozioökonomische Unterschiede [192, 193] wie auch die Wanderungsmotive [1] ausschlaggebend für Unterschiede im Fertilitätsverhalten zwischen Immigranten im Zielland sein. So setzen sich die Zuzüge in Deutschland aus verschiedenen Nationalitäten zusammen, die sich z. T. deutlich in der Fertilität unterscheiden. Während die Immigranten aus der Türkei eine relativ hohe Fertilität aufweisen, ist die Fertilität der Immigranten aus Polen, Rumänien, Ungarn oder auch Spanien z. T. so niedrig, dass

sie sogar die Fertilität der Einheimischen unterschreitet.<sup>11</sup> Darüber hinaus können die Unterschiede in der Fertilität zwischen Immigranten und Einheimischen durch Selektionsprozesse z. B. hinsichtlich des Alters, der Bildung und des Arbeitsplatzes nahezu vollständig erklärt werden [192, 193]. Außerdem wird die Fertilität der Migranten, die aufgrund einer Familienzusammenführung wandern, höher sein als von Migranten, die aufgrund einer Erwerbstätigkeit wandern. Folglich bestimmt die ethnische Zusammensetzung wie auch die Selektion hinsichtlich demographischer und sozioökonomischer Merkmale die Höhe der zu realisierenden Fertilität der Migranten (als eine gemeinsame Gruppe) in einem Zielland. Daneben haben auch die Bemühungen zur Integration (wie z. B. das Erlernen der Sprache im Zielland) oder aber auch die Wanderungspolitik als äußere Rahmenbedingung einen wesentlichen Einfluss auf das demographische Verhalten der Migranten im Zielland.

Der Aufbau des PPPM ermöglicht eine flexible Berücksichtigung all dieser Faktoren bei der Generierung der Annahmen zur Fertilität für die (direkten) Immi- und Emigranten und ihre jeweiligen Nachkommengenerationen. Für eine Bevölkerungsprognose mit dem PPPM ist die Annahme eines sukzessiven Anpassungsprozesses zwischen der Fertilität der Immigranten und Einheimischen über die Nachkommengenerationen zu empfehlen. In diesem Zusammenhang hat Sobotka 2008 [257] in einer vergleichenden Analyse der Fertilität von Immigranten und Einheimischen in 11 europäischen Ländern festgestellt, dass die TFR aller Immigranten zwischen 0,3 und 0,8 höher ist als bei den Einheimischen. Außerdem nimmt dieser Fertilitätsunterschied mit der Zeit zwischen den Immigranten und den Einheimischen ab. Ferner ist bei den Immigranten im Vergleich zu den Einheimischen das durchschnittliche Alter der Mütter bei Geburt niedriger, der Anteil der Kinderlosen geringer und der Anteil mit relativ vielen Kindern (bzw. einer höheren Parität) höher.

In Deutschland gibt es bislang keine Angaben zu den altersspezifischen Fertilitätsraten der direkten Zuzüge<sup>12</sup>. Ersatzweise können die altersspezifischen Fertilitätsraten ausländischer Frauen in Deutschland herangezogen werden. Allerdings unterscheiden sich diese beiden Personengruppen in einigen wesentlichen Punkten. So zählen zu den Ausländern ausschließlich Personen mit einer ausländischen Staatsangehörigkeit (und Staatenlose), wohingegen zu den

---

<sup>11</sup>So beträgt die TFR nach Daten von Eurostat im Jahr 2008 2,1 in der Türkei, 1,39 in Polen, 1,35 in Rumänien, 1,35 in Ungarn und 1,46 in Spanien.

<sup>12</sup>Allerdings verbessert sich die Datenlage zusehends: So wird z. B. im Mikrozensus 2008 der Migrationsstatus (wie auch die entsprechende Staatsangehörigkeit) neu erfasst, sodass umfassendere Auswertungen in Verbindung mit weiteren Variablen bzw. Merkmalen durchgeführt werden können [269].

direkten Zuzügen auch noch Personen mit einer deutschen Staatsangehörigkeit (wie z. B. deutsche Remigranten aus den USA) gehören [41]. Daneben kann sich die Aufenthaltsdauer zwischen den Ausländern und den direkten Zuzügen stark unterscheiden; während Ausländer im Durchschnitt (bereits) 18,2 Jahre (am Ende des Kalenderjahres 2008) in Deutschland sind [41], wandern direkte Zuzüge gerade (häufig) erst zu, sodass sich Ausländer im Gegensatz zu den direkten Zuzügen auch schon stärker an die Gegebenheiten (so z. B. auch in Bezug auf das fertile Verhalten im Sinne der Hypothese der Adaption) im Zielland angepasst haben können. Weiterhin wird die Fertilität der Ausländerinnen in Deutschland von der amtlichen Statistik unterschätzt. Dabei ist der erfasste Bevölkerungsbestand durch unterlassene Abmeldung von Remigrantinnen zu hoch, wohingegen die erfassten Geburten zu niedrig sind, da im Ausland stattfindende Geburten von in Deutschland gemeldeten ausländischen Frauen nicht in die Geburtenstatistik eingehen [249, 250]. Folglich können die von der amtlichen Statistik ausgewiesenen altersspezifischen Fertilitätsraten ausländischer Frauen als untere Schätzung für die altersspezifischen Fertilitätsraten der direkten Zuzüge dienen.

### **Mortalitätsprognose**

Empirische Studien haben gezeigt, dass die Sterblichkeit von Migranten im Zielland (respektive in Deutschland und Australien) oftmals geringer als in ihrem Herkunftsland ist [143, 178, 228, 229, 320]. Dieses Phänomen wird häufig mit dem so genannten *Healthy-Migrant Effekt* erklärt. Der Healthy-Migrant Effekt besagt, dass es sich bei Migranten um eine ausgewählte Gruppe von Personen handle, die, im Vergleich zu der restlichen, nicht wandernden Herkunftsbevölkerung, besonders fit bzw. gesund sei. Diese Fitness schlägt sich nicht nur in einem Sterblichkeitsvorteil der Migranten gegenüber ihrer Herkunftsbevölkerung, sondern auch, wie beispielsweise in Deutschland, gegenüber ihrer Zielbevölkerung nieder. Allerdings gehe, dem Healthy-Migrant Effekt zufolge, dieser Sterblichkeitsvorteil der Migranten gegenüber der Zielbevölkerung nach kurzer Zeit verloren, da die Integration beispielsweise im sozialen, ökonomischen und gesundheitlichen Bereich suboptimal sei und zu einer höheren Sterblichkeit der Migranten führe.

Diese Erklärung stimmt jedoch nicht mit der beobachtbaren Sterblichkeit der Migranten in Deutschland überein, nach der die Migranten über einen längeren Zeitraum eine deutlich geringere Sterblichkeit als die Einheimischen aufweisen. Jedoch wird die Sterblichkeit der Migranten in der amtlichen Statistik unterschätzt, da die erfassten Sterbefallzahlen zu niedrig und die erfassten Bevölkerungsbestände zu hoch sind. Dieser Effekt kann durch

unterlassene Abmeldungen und nicht gemeldete Sterbefälle von Zuwanderern im Ausland verursacht werden, sodass die tatsächliche Sterblichkeit der Zuwanderer in Deutschland etwas höher als die von der amtlichen Statistik erfasste Sterblichkeit der Migranten ist [142]. Allerdings zeigen auch Analysen zur Sterblichkeit von Zuwanderern in Deutschland mit anderen Datenquellen wie mit dem Ausländerzentralregister, dem Sozioökonomischen Panel oder auch mit Daten der deutschen Rentenversicherung einen Sterblichkeitsvorteil der Migranten gegenüber den Einheimischen der Zielbevölkerung in den ersten Jahren nach der Ankunft. Allerdings verschwindet dieser Sterblichkeitsvorteil mit der Länge des Aufenthalts der Migranten in der Zielbevölkerung zusehends [144, 229]. Razum [228] erklärt den Sterblichkeitsvorteil der direkten Zuzüge in Deutschland in den ersten Jahren damit, dass die direkten Zuzüge in Deutschland anderen Risikofaktoren ausgesetzt sind als in ihrem Herkunftsland. Dadurch können die direkten Zuzüge einerseits von der besseren medizinischen Versorgung im Zielland und andererseits von der erst ab dem Wanderungszeitpunkt einsetzenden, längeren Latenzzeit vor dem Ausbrechen nicht übertragbarer Krankheiten, denen sie nun im Zielland neu ausgesetzt sind, profitieren. Zusätzlich wird dieser Sterblichkeitsvorteil durch die Remigration kranker oder auch sozioökonomisch benachteiligter Migranten in ihr Herkunftsland verstärkt, da vermehrt gesunde Migranten im Zielland verbleiben. Allerdings nimmt dieser Sterblichkeitsvorteil der direkten Zuzüge mit zunehmender Länge ihres Aufenthalts im Zielland ab, da das Ende der seit der Ankunft im Zielland einsetzenden Latenzzeit neuer Krankheiten erreicht ist. Zudem können nun auch verstärkt Krankheiten bei den direkten Zuzügen im Gegensatz zu den Einheimischen auftreten, die durch die Risikofaktoren aus der Kindheit im Herkunftsland bedingt werden.<sup>13</sup> Daneben nehmen die Unterschiede in der Sterblichkeit zwischen den Einheimischen und jeder weiteren Nachkommengeneration der Migranten zunehmend ab, da sich die Lebensstile und damit auch die Risikofaktoren für Krankheiten zunehmend angleichen. Außerdem ist bei den Nachkommen auch die selektive Remigration kaum noch relevant.

Demzufolge ist für eine Bevölkerungsprognose mit dem PPPM die Modellierung eines Sterblichkeitsvorteils für die direkten Zuzüge und eines Annäherungsprozesses an die Sterblichkeit der Einheimischen für die entsprechenden Nachkommengenerationen bei der Generierung der Mortalitätsannahmen zu empfehlen. In diesem Zusammenhang hat Kohls [143] in einer empirischen Analyse die Sterblichkeit der ausländischen und einheimischen

---

<sup>13</sup>Ein Beispiel für eine erhöhte Sterblichkeit der Migranten gegenüber den Einheimischen ist Magenkrebs, der durch die Risikofaktoren in der Kindheit im Herkunftsland hervorgerufen werden kann.

Bevölkerung in Deutschland verglichen, wobei verschiedene Datenquellen zur Bestimmung der Sterblichkeit der ausländischen Bevölkerung (bzw. Migranten) verwendet worden sind. Den Daten der amtlichen Statistik zufolge beträgt der Unterschied in der Lebenserwartung bei Geburt zwischen der ausländischen und einheimischen Bevölkerung bei den Männern 6 und bei den Frauen 4 Jahre, wohingegen er den Daten des Ausländerzentralregisters zufolge bei den Männern 4 und bei den Frauen 2 Jahre beträgt. Demzufolge reduziert sich der Sterblichkeitsunterschied zwischen der ausländischen und einheimischen Bevölkerung, wenn Verzerrungen in den Daten eliminiert werden. Ferner ist die Sterblichkeit der ausländischen Bevölkerung in Deutschland in den jungen Altern (ca. bis zum Alter 15) etwas höher und in den darauffolgenden Altern niedriger als in der einheimischen Bevölkerung.<sup>14</sup>

In Deutschland gibt es keine Angaben zu den altersspezifischen Sterbewahrscheinlichkeiten der direkten Zuzüge. Ersatzweise können die altersspezifischen Sterbewahrscheinlichkeiten der ausländischen Bevölkerung herangezogen werden (vgl. Abschnitt 12.4.1 ab Seite 243). Allerdings unterscheiden sich diese beiden Personengruppen (wie bei der Fertilität) in einigen wesentlichen Merkmalen. So ist die Sterblichkeit der ausländischen Bevölkerung aufgrund der längeren Aufenthaltsdauer (im Sinne der Hypothese der Adaption und des Healthy-Migrant Effekts) vermutlich höher als die der direkten Zuzüge, zumal auch in Deutschland Geborene mit einer ausländischen Staatsangehörigkeit zur ausländischen Bevölkerung gehören. Allerdings kann dieser Effekt durch die selektive Remigration (gescheiterter Migranten) z. T. aufgehoben werden. Daneben ist die Sterblichkeit der direkten Zuzüge nicht soviel niedriger als die der ausländischen Bevölkerung anzusetzen, da deren von der amtlichen Statistik ausgewiesene Sterblichkeit durch unterlassene Abmeldung und nicht registrierte Sterbefälle im Ausland sowieso unterschätzt wird [142]. Folglich kann die von der amtlichen Statistik ausgewiesene altersspezifische Sterblichkeit der ausländischen Bevölkerung als eine (untere) Schätzung für die altersspezifische Sterblichkeit der direkten Zuzüge dienen.

---

<sup>14</sup>Die Sterblichkeit der Migranten ist nach den Daten des Ausländerzentralregisters (im Berichtsjahr 2006) in den Altern von 10 bis 64 um 10 Prozent höher als in der amtlichen Statistik; ab dem Alter 65 steigt dieser Sterblichkeitsunterschied sogar noch weiter an (bis zu 250 Prozent in der Altersgruppe 85+), sodass der Sterblichkeitsvorteil der Migranten gegenüber der einheimischen Bevölkerung insbesondere in den hohen Altern reduziert wird [143].

## 7.2. Schlussfolgerung — Vorschläge zur Generierung von Migrationsannahmen für eine Prognose mit dem PPPM

Dieses Kapitel gibt einen Überblick über die Vielzahl der bestehenden Ansätze zur Migrationsprognose, die verschiedene Informationsquellen nutzen und sich demzufolge für unterschiedlich lange Projektionshorizonte eignen. Generell können extrapolative von interpolativen, theoretisch erklärenden, bayesianischen und weiteren multikausalen Ansätzen unterschieden werden.

Die Prognose der Migration ist in der Vergangenheit häufig aufgrund kaum verfügbarer Daten oder/und einer (fälschlicherweise) unterstellten quantitativen Bedeutungslosigkeit vernachlässigt worden. Als Konsequenz ist die Migration entweder gar nicht oder oftmals nur vereinfachend mit einem konstanten Wert in die Zukunft extrapoliert worden. Allerdings kann die aus variierenden Wanderungsgründen stark schwankende Migrationsentwicklung mit dieser Vorgehensweise nur ungenügend eingefangen werden, sodass neben der vergangenen, rein quantitativen Entwicklung noch weitere Informationsquellen bei einer umfassenden Migrationsprognose berücksichtigt werden sollten. So können beispielsweise mit theoretisch erklärenden Ansätzen vielfältige Gründe für das Auftreten internationaler Migration aufgedeckt und in Zusammenhang mit Expertenmeinung oder/und entsprechenden mathematischen Modellen (auf der Basis empirischer Datenanalysen) prognostiziert werden. Multikausale Ansätze können dann quasi kulminierend (nahezu) alle verfügbaren Informationen und Methoden komplementär vereinen, um die Migration in all ihren potentiellen Ausprägungen (möglichst genau) zu prognostizieren.

Die Migration ist das Resultat eines komplexen Zusammenspiels diverser sozialer, ökonomischer und politisch(-rechtlich)er Faktoren im In- und Ausland, das von einem Prognostiker weder vollständig durchschaut noch exakt prognostiziert werden kann. Deshalb ist die Prognose der Migration besonders unsicher, sodass nicht nur ein, sondern mehrere Annahmenverläufe der Migration in einer Bevölkerungsprognose berücksichtigt werden sollten, um das potentielle Spektrum zukünftiger, mitunter stark heterogener Migrationsentwicklungen angemessen abdecken zu können.

Im PPPM können deshalb diverse Annahmenverläufe für die (totalen) Zu- und Fortzüge (nach Alter und Geschlecht) gewichtet in einer probabilistischen Bevölkerungsprognose



berücksichtigt werden, sodass die potentiellen Ausprägungen der Migration angemessen eingefangen werden können.

Daneben werden nicht nur die direkten Zu- und Fortzüge, sondern auch noch deren Nachkommengenerationen mit separaten Annahmen zur Fertilität und Mortalität (einzeln) prognostiziert, sodass einerseits sukzessive Anpassungsprozesse zwischen der Fertilität und Sterblichkeit der Migranten und Einheimischen (über die Nachkommen) modelliert und andererseits diverse langfristige bevölkerungsdynamische Effekte der Migranten auf die einheimische Bevölkerung detailliert analysiert werden können.

## 8. Bewertung von Bevölkerungsprognosen

Eine umfassende Bewertung von Bevölkerungsprognosen ist für deren Weiterentwicklung bzw. Verbesserung unerlässlich, da mit ihr überprüft werden kann, inwiefern die prognostizierten mit den tatsächlich eintreffenden Bevölkerungszahlen übereinstimmen und wodurch eventuelle Abweichungen bzw. Prognosefehler<sup>1</sup> entstehen.

Zur Bewertung von Bevölkerungsprognosen können prinzipiell diverse Gütekriterien,<sup>2</sup> die einzelne Arbeitsschritte eines Prognoseprozesses (vgl. Abschnitt 2.2 ab Seite 18) aus verschiedenen Perspektiven beleuchten, herangezogen werden. Grundsätzlich sollte ein Prognostiker mit einem Bevölkerungsprognosemodell die zukünftige Bevölkerungsentwicklung objektiv, zuverlässig, valide und genau prognostizieren. Das Gütekriterium der Objektivität misst dabei, ob die erzeugten Prognoseergebnisse unabhängig von den Einstellungen des jeweiligen Prognostikers (oder der Institution des Prognostikers) generiert worden sind. Daran schließt sich das Gütekriterium der Reliabilität, mit dem überprüft werden kann, ob sich die Prognoseergebnisse nach mehrmaliger Wiederholung verändern oder ob sie zuverlässig reproduzierbar sind. Mit dem Gütekriterium der Validität kann getestet werden, ob ein Bevölkerungsprognosemodell tatsächlich den vorher genau zu definierenden Prognosegegenstand misst. Darüber hinaus kann mit so genannten Fehlermaßen überprüft werden, wie genau die Prognoseergebnisse die tatsächlich eintreffenden Bevölkerungswerte wiedergeben. Ferner gibt es u. a. noch die Gütekriterien der Benutzerfreundlichkeit, der Transparenz oder auch der Entwicklungskosten. Die Vielseitigkeit der Gütekriterien ermöglicht deren sinnvollen Einsatz bei einer Fehleranalyse wie auch als Entscheidungshilfe für ein geeignetes Prognosemodell.

Der Fokus bei der Bewertung von Bevölkerungsprognosen richtet sich gemeinhin auf die

---

<sup>1</sup>Unter einem Prognosefehler ist die Abweichung der prognostizierten von den tatsächlich eingetroffenen Bevölkerungswerten zu verstehen.

<sup>2</sup>Eine Auflistung und Besprechung diverser Gütekriterien ist z. B. bei Diekmann [73], Smith et al. [256], Long [169], Ahlburg [2], Rogers [238] und Booth [30] zu finden.

Genauigkeit der Prognoseergebnisse. Deshalb werden nachfolgend Fehlermaße für deterministische Bevölkerungsprognosen kurz vorgestellt und anschließend hinsichtlich ihrer Anwendbarkeit für bzw. Übertragbarkeit auf probabilistische Bevölkerungsprognosen diskutiert. Abschließend sollen die wesentlichen Punkte der Diskussion über die Superiorität von simplen oder komplexen Bevölkerungsprognosemodellen erläutert werden, wobei Gütekriterien unterstützend eingesetzt werden können.

### 8.1. Fehlermaße

Es gibt eine Vielzahl bestehender Fehlermaße, die sich in ihrer Aussagekraft wie auch in dem, was sie eigentlich messen, deutlich voneinander unterscheiden.<sup>3</sup>

Fehlermaße können aufgrund ihrer unterschiedlichen Konstruktion für verschiedene Zwecke eingesetzt werden. So können sie z. B. neben dem Ausmaß und der Richtung der Prognosefehler auch die Sensitivität eines Bevölkerungsprognosemodells gegenüber bestimmten Parametervariationen messen. Je nach Fragestellung können dann z. B. Fehlermaße herangezogen werden, die einfache oder mittlere Fehler über ein oder mehrere Projektionsintervalle messen bzw. äußerst sensibel oder robust auf Ausreißer reagieren. Demzufolge sollte die Wahl eines Fehlermaßes immer auf die jeweilige Fragestellung abgestimmt werden, um eine valide Bewertung zu erhalten [12, 14, 87, 125, 126].

Herkömmliche Fehlermaße sind vornehmlich für deterministische Bevölkerungsprognosen konzipiert, sodass sie entsprechend erweitert werden müssen, um auch für probabilistische Bevölkerungsprognosen angewendet werden zu können.

---

<sup>3</sup>In der deutschen Literatur sind Fehlermaße weniger verbreitet. Jedoch hat sich z. B. Bretz [40] mit der Treffsicherheit früherer Bevölkerungsprognosen beschäftigt, indem er die fortgeschriebene mit der tatsächlich eingetroffenen Entwicklung der Bevölkerung und einzelner demographischer Komponenten miteinander verglichen hat.

### 8.1.1. Herkömmliche Fehlermaße für deterministische Bevölkerungsprognosen

Hyndman und Koehler [117]<sup>4</sup> unterscheiden skalenabhängige von skalenunabhängigen Fehlermaßen und Fehlermaße aus relativen Fehlern von relativen Fehlermaßen. Skalenunabhängige Fehlermaße ermöglichen einen Vergleich der Prognosefehler unterschiedlich großer Bevölkerungen, wohingegen die Höhe skalenabhängiger Fehlermaße von der Größe einer zu prognostizierenden Bevölkerung abhängt. Um die Prognosekraft verschiedener Prognosemethoden zu vergleichen, können Fehlermaße aus relativen Fehlern und relative Fehlermaße herangezogen werden. Fehlermaße aus relativen Fehlern setzen Prognosefehler einer formalen und einer naiven Prognose<sup>5</sup> für einzelne Projektionsintervalle ins Verhältnis, wohingegen relative Fehlermaße Lagemaße wie mittlere Fehler einer formalen und einer naiven Prognose über mehrere Projektionsintervalle ins Verhältnis setzen.

Im Folgenden sollen einige Fehlermaße für jede dieser vier Kategorien beispielhaft erläutert werden. In den Berechnungsformeln für die einzelnen Fehlermaße wird eine einheitliche Notation verwendet, in der  $F$  für den prognostizierten und  $A$  für den tatsächlich eingetroffenen Bevölkerungswert in einem Projektionsintervall  $t$  steht. Darüber hinaus umfasst der Projektionshorizont insgesamt  $N$  Projektionsintervalle, wobei  $n$  das letzte zu betrachtende Projektionsintervall einer Prognose kennzeichnet.

#### Beispiele für skalenabhängige Fehlermaße

**Prognosefehler — E — Error** Der einfache Prognosefehler  $E$  in einem Projektionsintervall  $t$  ist die Differenz aus dem prognostizierten und dem tatsächlich eingetroffenen Bevölkerungswert in diesem Projektionsintervall (vgl. Formel 8.1).

$$E_t = F_t - A_t \tag{8.1}$$

Der Prognosefehler  $E$  kann sowohl positive als auch negative Werte annehmen. Bei einem positiven Prognosefehler liegt der prognostizierte über dem tatsächlich eingetroffenen Wert, wohingegen bei einem negativen Prognosefehler der prognostizierte unter dem tatsächlich

---

<sup>4</sup>Außer bei Hyndman und Koehler [117] werden Fehlermaße beispielsweise auch bei Armstrong [12], Smith et al. [256], Keilman und Pham [130], Keilman [127] und Pflaumer [218] näher erläutert und vielfältig angewendet.

<sup>5</sup>Unter einer naiven Bevölkerungsprognose ist beispielsweise die Annahme konstanter, zuletzt beobachteter Bevölkerungswerte zu verstehen.

eingetroffenen Wert liegt. Der einfache Prognosefehler beträgt *null*, wenn der prognostizierte und der reale Wert übereinstimmen. Die Höhe des Prognosefehlers richtet sich nach der Höhe der Prognoseergebnisse, sodass er skalenabhängig ist.

Es können nun diverse weitere skalenabhängige Fehlermaße berechnet werden, die die Verteilung des einfachen Prognosefehlers über mehrere Projektionsintervalle charakterisieren.

**Mittlerer Prognosefehler — ME — Mean Error** Der mittlere Prognosefehler *ME* bestimmt sich über die Summe der Prognosefehler *E* über alle Projektionsintervalle *t* geteilt durch die Anzahl aller Projektionsintervalle *N* (vgl. Formel 8.2).

$$ME = \frac{\sum_{t=1}^n E_t}{N} \quad (8.2)$$

Der mittlere Prognosefehler *ME* ist ein Lagemaß, das den arithmetischen Mittelwert der Prognosefehler *E* einer Prognose über alle Projektionsintervalle bestimmt. Er kann sowohl positive als auch negative Werte annehmen. Ist der mittlere Prognosefehler positiv, so überschätzen die prognostizierten Werte im Durchschnitt die realen Werte. Hingegen unterschätzen die prognostizierten Werte im Durchschnitt die realen Werte, wenn der mittlere Prognosefehler negativ ist. Demnach zeigt der *ME* die Richtung der Prognosefehler an. Stimmen alle prognostizierten Werte exakt mit den realen Werten überein, so nimmt der mittlere Prognosefehler den Wert *null* an. Darüber hinaus kann der mittlere Prognosefehler auch *null* sein, wenn sich positive und negative Prognosefehler entsprechend ausgleichen.

Anstatt des arithmetischen Mittels kann auch der Median der einfachen Prognosefehler über den Projektionshorizont bestimmt werden. Dieses Fehlermaß wird dann als Median der Prognosefehler, *MdE* (Median Error), bezeichnet. Der *MdE* kann sowohl positive als auch negative Werte annehmen. Der *MdE* nimmt den Wert *null* an, wenn mehr als die Hälfte aller prognostizierten Werte mit den entsprechenden realen Werten übereinstimmen. Ist der *MdE* größer bzw. kleiner als der *ME*, so ist die Verteilung der Prognosefehler linksschief bzw. rechtsschief. Der *MdE* ist gegenüber Ausreißern und Fehlern in den Daten weniger sensibel als der *ME*.

**Mittlerer absoluter Prognosefehler — MAE — Mean Absolute Error** Der mittlere absolute Prognosefehler *MAE* bestimmt sich über die Summe der absoluten Prognosefehler  $|E|$  über alle Projektionsintervalle *t* geteilt durch die Anzahl aller Projektionsintervalle *N*

(vgl. Formel 8.3).

$$MAE = \frac{\sum_{t=1}^n |E_t|}{N} \quad (8.3)$$

Der mittlere absolute Prognosefehler  $MAE$  ist ein Lagemaß, das den arithmetischen Mittelwert der absoluten Prognosefehler  $|E|$  über den Projektionshorizont bestimmt. Im Gegensatz zum mittleren Prognosefehler  $ME$  kann der mittlere absolute Prognosefehler aufgrund der Betragsbildung ausschließlich positive Werte annehmen, sodass sich positive und negative Prognosefehler nicht mehr ausgleichen können. Der mittlere absolute Prognosefehler hat einen Minimalwert von *null*, wenn die prognostizierten mit den realen Werten übereinstimmen. Je höher der Wert des mittleren absoluten Prognosefehlers ist, desto größer sind die (positiven und negativen) Abweichungen der prognostizierten von den realen Werten. Demnach bestimmt der  $MAE$  das Ausmaß und nicht wie der  $ME$  die Richtung der Prognosefehler.

Statt des arithmetischen Mittelwerts kann auch der Median der absoluten Prognosefehler,  $MdAE$  (Median Absolute Error), über den Projektionshorizont bestimmt werden. Die Höhe des  $MdAE$  richtet sich nach der Höhe der Werte der absoluten Prognosefehler. Da die absoluten Prognosefehler aufgrund der Betragsbildung durchweg positiv sind, nimmt auch der Median der absoluten Prognosefehler ausschließlich positive Werte an. Der  $MdAE$  entspricht seinem Minimalwert *null*, wenn mehr als die Hälfte aller prognostizierten Werte mit den jeweiligen realen Werten übereinstimmt. Ist der Median größer bzw. kleiner als der arithmetische Mittelwert der absoluten Prognosefehler, so ist die Verteilung der absoluten Prognosefehler linksschief bzw. rechtsschief. Der  $MdAE$  ist robuster gegenüber Ausreißern als der  $MAE$ .

**Mittlerer quadrierter Prognosefehler — MSE — Mean Squared Error** Der mittlere quadrierte Prognosefehler  $MSE$  berechnet sich aus der Summe der quadrierten Prognosefehler  $E$  über den Projektionshorizont geteilt durch die Anzahl aller Projektionsintervalle  $N$  (vgl. Formel 8.4).

$$MSE = \frac{\sum_{t=1}^n E_t^2}{N} \quad (8.4)$$

Der mittlere quadrierte Prognosefehler  $MSE$  ist ein Lagemaß, das die Verteilung der quadrierten Prognosefehler charakterisiert. Prinzipiell entspricht der  $MSE$  der Varianz der Prognosefehler um den jeweils tatsächlich eingetroffenen Wert. Durch die Quadrierung der Prognosefehler kann der  $MSE$  ausschließlich positive Werte annehmen. Der Minimalwert

des  $MSE$  von  $null$  wird genau dann angenommen, wenn alle prognostizierten mit den entsprechenden realen Werten übereinstimmen. Je höher der Wert des  $MSE$  ist, desto größer sind die Abweichungen der prognostizierten von den realen Werten. Demzufolge kann auch mit dem  $MSE$  (wie mit dem  $MAE$ ) das Ausmaß und nicht die Richtung der Prognosefehler bestimmt werden. Der  $MSE$  reagiert besonders sensibel auf Ausreißer, da hohe Prognosefehler, ob nun im negativen oder positiven Bereich, deutlich stärker ins Gewicht fallen als niedrigere Prognosefehler.

Anstatt des arithmetischen Mittelwerts kann auch der Median der quadrierten Prognosefehler mit dem  $MdSE$  (Median Squared Error) bestimmt werden. Der  $MdSE$  ist robuster gegenüber Ausreißern als der  $MSE$ .

**Wurzel aus dem mittleren quadrierten Prognosefehler — RMSE — Root Mean Squared Error** Der  $RMSE$  berechnet sich aus der Wurzel des mittleren quadrierten Prognosefehlers  $MSE$  (vgl. Formel 8.5).

$$RMSE = \sqrt{\frac{\sum_{t=1}^n E_t^2}{N}} \quad (8.5)$$

Prinzipiell entspricht der  $RMSE$  der Standardabweichung der Prognosefehler um den jeweils tatsächlich eingetroffenen Wert. Der  $RMSE$  kann ausschließlich positive Werte annehmen. Je höher der Wert des  $RMSE$  ist, desto größer sind die Prognosefehler. Der Minimalwert von  $null$  tritt dann beim  $RMSE$  ein, wenn alle prognostizierten Werte mit den entsprechenden realen Werten übereinstimmen. Genau wie der  $MSE$  reagiert auch der  $RMSE$  sehr sensibel auf Ausreißer, da höhere Prognosefehler durch die Quadrierung ein deutlich höheres Gewicht erhalten. Dennoch ist der  $RMSE$  in der Vergangenheit eines der am häufigsten verwendeten Fehlermaße zur Bewertung von Prognoseergebnissen gewesen [12].

Neben dem  $RMSE$  gibt es auch den  $RMdSE$  (Root Median Squared Error), der die Wurzel aus dem Median aller quadrierten Prognosefehler bildet. Der  $RMdSE$  ist robuster gegenüber Ausreißern als der  $RMSE$ .

**Zusammenfassung** Skalenabhängige Fehlermaße basieren auf dem einfachen Prognosefehler  $E$ . Die Verteilung dieser z. T. umgeformten Prognosefehler ( $E$ ,  $AE$ ,  $SE$ ) über den Projektionshorizont kann mit diversen (Lage)Fehlermaßen charakterisiert werden. Die Fehlermaße messen entweder die Genauigkeit (accuracy) oder systematische Verzerrungen

(bias) der Prognoseergebnisse. Darüber hinaus unterscheiden sie sich in ihrer Sensibilität gegenüber Ausreißern.

Die Höhe (bzw. Skala) des Prognosefehlers  $E$  (und seiner Varianten) ist von der Höhe (bzw. Skala) der Prognoseergebnisse abhängig ist. Demzufolge ist ein Vergleich der Prognosefehler unterschiedlich großer Bevölkerungen mit skalenabhängigen Fehlermaßen nicht möglich. Im Gegensatz dazu können mit skalenunabhängigen Fehlermaßen Prognosefehler unterschiedlich großer Bevölkerungen miteinander verglichen werden, da die Abweichungen zur Bevölkerungsgröße ins Verhältnis gesetzt werden.

### Beispiele für skalenunabhängige Fehlermaße

**Prozentualer Prognosefehler — PE — Percentage Error** Der prozentuale Prognosefehler  $PE$  in einem Projektionsintervall  $t$  berechnet sich aus der Differenz des prognostizierten und realen Bevölkerungswerts geteilt durch den realen Bevölkerungswert im Projektionsintervall  $t$  (vgl. Formel 8.6).

$$PE_t = \frac{F_t - A_t}{A_t} \quad (8.6)$$

Der prozentuale Prognosefehler,  $PE$ , gibt den Anteil der Abweichung des prognostizierten vom tatsächlich beobachteten Bevölkerungswert relativ zum realen Bevölkerungswert an. Dadurch ist der prozentuale Prognosefehler skalenunabhängig. Der prozentuale Prognosefehler kann sowohl positive als auch negative Werte annehmen. Bei einem positiven prozentualen Prognosefehler ist der prognostizierte Wert größer als der reale Wert; wohingegen bei einem negativen prozentualen Prognosefehler der reale Wert größer als der prognostizierte Wert ist. Stimmen der prognostizierte und der reale Wert überein, so entspricht der prozentuale Prognosefehler *null*. Eine generelle Schwäche des prozentualen Prognosefehlers liegt in seiner Anfälligkeit gegenüber kleinen realen Werten nahe oder gleich *null*. In diesem Fall droht der prozentuale Prognosefehler schnell unverhältnismäßig groß zu werden oder bei einer Division durch *null* nicht definiert zu sein.

**(Lage)Fehlermaße auf der Basis des prozentualen Prognosefehlers** Analog zum einfachen Prognosefehler  $E$  kann auch die Verteilung der (z. T. umgeformten) prozentualen Prognosefehler ( $PE$ ,  $APE$  (Absolute Percentage Error),  $SPE$  (Squared Percentage Error)) über den Projektionshorizont mit entsprechenden (Lage)Fehlermaßen charakterisiert werden. Die Berechnung entspricht den (Lage)Fehlermaßen mit dem einfachen Prognosefehler, nur



dass nun der prozentuale anstatt des einfachen Prognosefehlers verwendet wird.

Das Ausmaß bzw. die Genauigkeit der Prognosefehler kann z. B. mit dem *MAPE* (Mean Absolute Percentage Error), dem *MdAPE* (Median Absolute Percentage Error), dem *MSPE* (Mean Squared Percentage Error), dem *MdSPE* (Median Squared Percentage Error), dem *RMSPE* (Root Mean Squared Percentage Error) oder auch mit dem *RMdSPE* (Root Median Squared Percentage Error) gemessen werden. Auch diese Fehlermaße unterscheiden sich in ihrer Sensibilität gegenüber Ausreißern. So sind Fehlermaße, die die Verteilung der prozentualen Prognosefehler über den Projektionshorizont mit einem Median charakterisieren, robuster als Fehlermaße, die dafür einen arithmetischen Mittelwert berechnen. Darüber hinaus erhalten Ausreißer bei Fehlermaßen, die den quadrierten prozentualen Prognosefehler verwenden, ein höheres Gewicht als bei Fehlermaßen, die den einfachen prozentualen Prognosefehler verwenden.

Die Richtung bzw. die systematische Verzerrung (bias) der Prognosefehler kann z. B. mit dem *MALPE* (Mean Algebraic Percentage Error) oder auch mit dem *MdPE* (Median Percentage Error) gemessen werden. Dabei reagiert der *MdPE* robuster auf Ausreißer als der *MALPE*.

Fehlermaße können nicht nur eine Prognosemethode, sondern gleich zwei Prognosemethoden vergleichend auf ihre Prognosekraft untersuchen bzw. bewerten. Dazu können Fehlermaße aus relativen Fehlern und relative Fehlermaße herangezogen werden.

### Beispiele für Fehlermaße aus relativen Fehlern

**Relativer absoluter Prognosefehler — RAE — Relative Absolute Error** Der relative absolute Prognosefehler *RAE* in einem Projektionsintervall  $t$  ist der Quotient aus dem einfachen Prognosefehler einer formalen und einer naiven Prognosemethode (vgl. Formel 8.7).

$$RAE_t = \frac{|E_t|}{|E_t^{naiv}|} \quad (8.7)$$

Der relative absolute Prognosefehler, *RAE*, vergleicht die Prognosefehler einer formalen und einer naiven Prognosemethode für jedes Projektionsintervall  $t$  separat. Der *RAE* kann ausschließlich positive Werte annehmen. Entspricht der *RAE* exakt dem Wert *null*, so beträgt der Prognosefehler der formalen Prognose ebenfalls *null*. Liegt der *RAE* über *null* und unter *eins*, so ist der Prognosefehler der formalen Prognose kleiner als der der naiven Prognose.

Ist der  $RAE$  größer als *eins*, so ist der Prognosefehler der formalen Prognose größer als der der naiven Prognose. Beträgt der  $RAE$  exakt *eins*, so sind die Prognosefehler der beiden zu vergleichenden Prognosen identisch. Problematisch ist die Berechnung des  $RAE$ , wenn der Prognosefehler der naiven Prognose *null* oder sehr klein ist, da dann der  $RAE$  nicht definiert oder unverhältnismäßig groß ist. Armstrong [12] schlägt deshalb die Berechnung eines *winsorisierten*  $RAE$  vor, bei dem extrem große und extrem kleine Werte abgeschnitten werden (vgl. Formel 8.8).

$$WRAE_t = \begin{cases} 0,01 & \text{wenn } RAE_t < 0,01 \\ RAE_t & \text{wenn } 0,01 \leq RAE_t \leq 10 \\ 10 & \text{wenn } RAE_t > 10 \end{cases} \quad (8.8)$$

Bei einem Fehlermaß aus relativen Fehlern können anstatt der einfachen Prognosefehler auch die prozentualen Prognosefehler einer formalen und einer naiven Prognose ins Verhältnis gesetzt werden. Der  $RAE$  und der  $RPE$  (bzw. der  $WRAE$  und der  $WRPE$ ) sind identisch, wenn sie bei denselben Prognoseergebnissen angewendet werden.

**(Lage)Fehlermaße auf der Basis des relativen absoluten Prognosefehlers** Die Verteilung der relativen absoluten Fehler über den Projektionshorizont kann mit diversen (Lage)Fehlermaßen charakterisiert werden. So können beispielsweise der  $MRAE$  (Mean Relative Absolute Error), der  $MdRAE$  (Median Relative Absolute Error) oder auch der  $GMRAE$  (Geometric Mean Relative Absolute Error) zum Vergleich der Genauigkeit zweier verschiedener Prognosemethoden berechnet werden. Armstrong und Collopy [14] empfehlen den  $MdRAE$  für den Vergleich verschiedener Prognosemethoden aufgrund seiner Robustheit gegenüber Ausreißern.

### Beispiele für relative Fehlermaße

Relative Fehlermaße vergleichen die Prognosekraft zweier Prognosemethoden, indem sie das gleiche (Lage)Fehlermaß einer formalen und einer naiven Prognose ins Verhältnis setzen. Dazu können alle skalenabhängigen und skalenunabhängigen (Lage)Fehlermaße verwendet werden. Beispielhaft soll nachfolgend der relative mittlere Prognosefehler  $RelMAE$  genauer erläutert werden.

**Relativer mittlerer Prognosefehler — RelMAE — Relative Mean Error** Der relative mittlere Prognosefehler  $RelMAE$  entspricht dem Quotient aus dem mittleren absoluten Prognosefehler einer formalen und einer naiven Prognose (vgl. Formel 8.9).

$$RelMAE = \frac{MAE}{MAE^{naiv}} \quad (8.9)$$

Der relative mittlere Prognosefehler vergleicht die mittleren absoluten Prognosefehler einer formalen und einer naiven Prognose über den gesamten Projektionshorizont. Der *RelMAE* kann ausschließlich positive Werte annehmen. Die *MAE* der formalen und naiven Prognose sind identisch, wenn der *RelMAE* den Wert *eins* annimmt. Liegt der *RelMAE* zwischen *null* und *eins*, so ist der *MAE* der naiven Prognose größer als der der formalen Prognose. Ist der *RelMAE* hingegen größer als *eins*, so ist der *MAE* der formalen Prognose größer als der der naiven Prognose.

### Zusammenfassung

Die in diesem Abschnitt aufgeführten Fehlermaße geben einen Überblick über die Vielfalt der bestehenden Fehlermaße, jedoch stellen sie lediglich eine Auswahl dar.

Skalenabhängige und skalenunabhängige Fehlermaße messen die Prognosefehler einer Prognosemethode, wohingegen die Fehlermaße aus relativen Fehlern und die relativen Fehlermaße die Prognosefehler zweier verschiedener Prognosemethoden miteinander vergleichen. Darüber hinaus bestimmt die Konstruktion der Fehlermaße, ob sie das Ausmaß oder die Richtung der Prognosefehler messen und ob sie dabei sensibel oder robust auf Ausreißer reagieren. Demzufolge sollte bei der Bewertung einer Bevölkerungsprognose ein Fehlermaß verwendet werden, dass den tatsächlichen Anforderungen entspricht. Wenn z. B. ein Bevölkerungsprognosemodell kalibriert werden soll, ist es entscheidend, dass ein Fehlermaß auf kleine Veränderungen sensibel reagiert. Sollen aber z. B. zwei Prognosemethoden miteinander verglichen werden, so sollte ein Fehlermaß robuster gegenüber Ausreißern, die z. B. auch durch Fehler in den Daten hervorgerufen werden können, reagieren.

Herkömmliche Fehlermaße sind generell zur Messung der Prognosefehler deterministischer Bevölkerungsprognosen konzipiert. Allerdings sollten auch die Ergebnisse probabilistischer Bevölkerungsprognosen mit Fehlermaßen bewertet werden können.

#### 8.1.2. Fehlermaße für probabilistische Bevölkerungsprognosen

Probabilistische Bevölkerungsprognosen prognostizieren die zukünftige Bevölkerungsentwicklung mit einem Unsicherheitsbereich aus vornehmlich zwei Gründen. Erstens kann von

einem Prognostiker nicht das komplexe Zusammenspiel aller wesentlichen Determinanten durchschaut und vorhergesagt werden. Zweitens ist die zukünftige Entwicklung grundsätzlich nicht vorherbestimmt bzw. deterministisch, sondern chaotisch bzw. stochastisch.

Dies wirft die Frage auf, inwiefern eine probabilistische Bevölkerungsprognose überhaupt bewertet werden kann. Im Prinzip kann eine probabilistische Bevölkerungsprognose nur dahingehend bewertet werden, ob die tatsächliche zukünftige Bevölkerungsentwicklung prognostiziert worden ist. Allerdings kann daraus kein Rückschluss auf die Qualität der Annahmen einer Prognose erfolgen, da die *wahre* Wahrscheinlichkeit der potentiellen Ereignisse, die in der Zukunft stattfinden können, unbekannt ist.<sup>6</sup>

Folglich können Fehlermaße lediglich die prognostizierte und die reale Bevölkerungsentwicklung im Nachhinein vergleichen. Dabei kann überprüft werden, ob die realen Bevölkerungswerte in der Ergebnisverteilung enthalten sind, in welchem Ausmaß und in welcher Richtung die prognostizierten um die realen Werte streuen und ob der Verlauf der realen und der prognostizierten Bevölkerungsentwicklung (un)ähnlich ist.

Fehlermaße zur Bewertung probabilistischer Bevölkerungsprognosen sollten daher die Breite und die Form der Ergebnisverteilung im Querschnitt (für ein Projektionsintervall) wie auch im Längsschnitt (über den Projektionshorizont) im Hinblick auf die tatsächlich eingetroffene Bevölkerungsentwicklung evaluieren. Nachfolgend werden deshalb einige in dieser Arbeit zu entwickelnde Fehlermaße zur Bewertung probabilistischer Bevölkerungsprognosen vorgestellt, die ein oder mehrere dieser Bewertungskriterien aufgreifen.

Die Fehlermaße zur Bewertung probabilistischer Bevölkerungsprognosen lehnen sich dabei an die Fehlermaße zur Bewertung deterministischer Bevölkerungsprognosen an. Um nun aber mehrere Ergebnisverläufe einer Ergebnisverteilung mit nur einem Fehlermaß bewerten zu können, müssen zusätzliche Variablen eingeführt werden. So gibt  $R$  die Anzahl aller zu

---

<sup>6</sup>So kann zwar z. B. die Wahrscheinlichkeitsverteilung der getroffenen Annahmen mit der *wahren* Wahrscheinlichkeitsverteilung übereinstimmen, allerdings würde bei Eintritt eines weniger wahrscheinlichen Ereignisses die Prognose fälschlicherweise als ungenauer eingestuft werden als wenn das Ereignis mit der höchsten Wahrscheinlichkeit eingetroffen wäre. Umgekehrt würde eine Prognose als genauer eingestuft werden, deren Wahrscheinlichkeitsverteilung zwar nicht mit der *wahren* Wahrscheinlichkeitsverteilung übereinstimmt, aber die das eigentlich unwahrscheinliche, aber dennoch eintretende Ereignis als sehr wahrscheinlich einstuft.

evaluierenden Ergebnisverläufe einer Ergebnisverteilung an. Darüber hinaus erhält jeder der in aufsteigender Reihenfolge sortierten, zu evaluierenden Ergebnisverläufe eine Nummer  $i$ , wobei der niedrigste Ergebnisverlauf die Nummer  $i = 0$  erhält. Damit kann jedem Ergebnisverlauf  $i$  ein Quantil der Ordnung  $p$  der Ergebnisverteilung,  $p_i$ , zugeordnet werden (vgl. Formel 8.10).

$$p_i = \frac{i}{R-1} \quad (8.10)$$

### **Mittlerer Prognosefehler einer Ergebnisverteilung — MERD — Mean Error of a Result Distribution**

Der mittlere Prognosefehler einer Ergebnisverteilung  $MERD$  in einem Projektionsintervall  $t$  entspricht der Summe der Abweichungen aller prognostizierten von den realen Bevölkerungswerten geteilt durch die Anzahl aller generierten Ergebnisverläufe (vgl. Formel 8.11 auf Seite 157).

$$MERD_t = \frac{1}{R} \sum_i F_{p_i,t} - A_t \quad (8.11)$$

Der mittlere Prognosefehler einer Ergebnisverteilung,  $MERD$ , gibt somit die durchschnittliche Abweichung aller prognostizierten Werte vom realen Bevölkerungswert in einem Projektionsintervall  $t$  an. Der  $MERD$  kann sowohl positive als auch negative Werte annehmen. Ist der  $MERD$  negativ, so unterschätzen die prognostizierten Bevölkerungswerte im Durchschnitt den realen Wert. Entspricht der  $MERD$  exakt dem Wert *null*, so gleichen sich die positiven und negativen Abweichungen im Durchschnitt in der Höhe aus. Ist der  $MERD$  hingegen positiv, überschätzen die prognostizierten Bevölkerungswerte im Durchschnitt den realen Wert. Demzufolge kann mit dem mittleren Prognosefehler einer Ergebnisverteilung,  $MERD$ , die Richtung des Prognosefehlers einer Ergebnisverteilung bestimmt werden. Darüber hinaus ist der  $MERD$  skalenabhängig, sodass er nicht zum Vergleich der Prognosefehler unterschiedlich großer Bevölkerungen herangezogen werden kann.

### **Mittlerer prozentualer Prognosefehler einer Ergebnisverteilung — MALPERD — Mean Algebraic Percentage Error of a Result Distribution**

Der mittlere prozentuale Prognosefehler einer Ergebnisverteilung,  $MALPERD$ , in einem Projektionsintervall  $t$  berechnet sich über die Summe der Quotienten aus den Abweichungen der

prognostizierten von dem realen Bevölkerungswert in Relation zum realen Bevölkerungswert geteilt durch die Anzahl aller prognostizierten Bevölkerungswerte (vgl. Formel 8.12).

$$MALPERD_t = \frac{1}{R} \sum_i \frac{F_{p_i,t} - A_t}{A_t} \quad (8.12)$$

Der mittlere prozentuale Prognosefehler einer Ergebnisverteilung, *MALPERD*, gibt die durchschnittliche relative Abweichung der prognostizierten von dem realen Bevölkerungswert im Verhältnis zum realen Bevölkerungswert in einem Projektionsintervall  $t$  an. Der *MALPERD* kann sowohl positive als auch negative Werte annehmen. Ist der *MALPERD* negativ bzw. positiv, so unter- bzw. überschätzen die prognostizierten den realen Bevölkerungswert im Durchschnitt. Entspricht der *MALPERD* exakt dem Wert *null*, so sind die positiven und die negativen Abweichungen im Durchschnitt in ihrer Höhe identisch. Folglich kann mit dem *MALPERD* (wie auch mit dem *MERD*) die Richtung des Prognosefehlers einer Ergebnisverteilung bestimmt werden. Daneben ist der *MALPERD* skalenunabhängig, sodass mit ihm die Prognosefehler von Ergebnisverteilungen unterschiedlich großer Bevölkerungen miteinander verglichen werden können.

#### **Mittlerer absoluter Prognosefehler einer Ergebnisverteilung — MAERD — Mean Absolute Error of a Result Distribution**

Der mittlere absolute Prognosefehler einer Ergebnisverteilung, *MAERD*, in einem Projektionsintervall  $t$  wird über die Summe der absoluten Abweichungen der prognostizierten von dem realen Bevölkerungswert geteilt durch die Anzahl aller prognostizierten Bevölkerungswerte ermittelt (vgl. Formel 8.13).

$$MAERD_t = \frac{1}{R} \sum_i |F_{p_i,t} - A_t| \quad (8.13)$$

Der mittlere absolute Prognosefehler einer Ergebnisverteilung, *MAERD*, gibt die durchschnittliche absolute Abweichung der prognostizierten von dem realen Bevölkerungswert in einem Projektionsintervall  $t$  an. Der *MAERD* kann durch die Betragsbildung ausschließlich positive Werte annehmen. Je höher der *MAERD* ist, desto größer sind auch die Abweichungen bzw. der Prognosefehler der Ergebnisverteilung. Demnach kann mit dem *MAERD* nicht die Richtung, sondern das Ausmaß des Prognosefehlers der Ergebnisverteilung bestimmt werden. Der *MAERD* ist wie der *MERD* skalenabhängig.

### Mittlerer absoluter prozentualer Prognosefehler einer Ergebnisverteilung — MAPERD — Mean Absolute Percentage Error of a Result Distribution

Der mittlere absolute prozentuale Prognosefehler einer Ergebnisverteilung, *MAPERD*, in einem Projektionsintervall  $t$  berechnet sich über die Summe der Quotienten aus den absoluten Abweichungen der prognostizierten vom realen Bevölkerungswert in Relation zum realen Bevölkerungswert geteilt durch die Anzahl aller prognostizierten Bevölkerungswerte (vgl. Formel 8.14).

$$MAPERD_t = \frac{1}{R} \sum_i \left| \frac{F_{p_i,t} - A_t}{A_t} \right| \quad (8.14)$$

Der mittlere absolute prozentuale Prognosefehler einer Ergebnisverteilung, *MAPERD*, gibt den durchschnittlichen Anteil der absoluten Abweichungen der prognostizierten vom realen Bevölkerungswert relativ zum realen Bevölkerungswert in einem Projektionsintervall  $t$  an. Der *MAPERD* kann durch die Betragsbildung ausschließlich positive Werte annehmen. Je höher der *MAPERD*, desto höher sind auch die Abweichungen der prognostizierten vom realen Bevölkerungswert. Mit dem *MAPERD* kann das Ausmaß des Prognosefehlers einer Ergebnisverteilung gemessen werden. Zudem ist der *MAPERD* skalenunabhängig.

### Lokalisation des realen Werts in der Ergebnisverteilung — LARD — Location of Actual value in Result Distribution

Die Lokalisation des realen Werts in der Ergebnisverteilung, *LARD*, bestimmt sich aus der Abweichung des niedrigsten prognostizierten vom realen Bevölkerungswert geteilt durch die Differenz aus dem höchsten und dem niedrigsten prognostizierten Bevölkerungswert in einem Projektionsintervall  $t$  (vgl. Formel 8.15).

$$LARD_t = \frac{A_t - F_{p_i=0,t}}{F_{p_i=1,t} - F_{p_i=0,t}} \quad (8.15)$$

*LARD* gibt somit die Position des tatsächlich eintreffenden Bevölkerungswerts  $A$  zwischen dem oberen und unteren Grenzverlauf der Ergebnisverteilung für ein Projektionsintervall  $t$  an. *LARD* kann sowohl positive als auch negative Werte annehmen. Liegt *LARD* zwischen *null* und *eins*, so liegt der reale Wert innerhalb der Ergebnisverteilung. Entspricht *LARD* exakt dem Wert *null* bzw. *eins*, so entspricht der reale Bevölkerungswert dem unteren bzw. oberen Grenzwert der Ergebnisverteilung. Ist *LARD* kleiner als *null* bzw. größer als *eins*, so liegt der reale Bevölkerungswert unterhalb bzw. oberhalb der Ergebnisverteilung.

### Weitere Fehlermaße zur Bewertung probabilistischer Bevölkerungsprognosen

Die bisher vorgestellten Fehlermaße zur Bewertung probabilistischer Bevölkerungsprognosen können nicht nur für ein Projektionsintervall  $t$ , sondern auch zusätzlich für den gesamten Projektionshorizont berechnet werden. Dazu kann beispielsweise der Verlauf der Fehlermaßwerte über den Projektionshorizont analysiert oder auch ein Lagemaß (wie z.B. der arithmetische Mittelwert) der Fehlermaßwerte über alle Projektionsintervalle bestimmt werden.

Außerdem gibt es noch zahlreiche Möglichkeiten, die vorgestellten Fehlermaße zur Bewertung probabilistischer Bevölkerungsprognosen zu verändern. So können z.B. besonders hohe Ausreißer stärker bei der Berechnung des Prognosefehlers ins Gewicht fallen, wenn die Abweichungen der prognostizierten von dem jeweils realen Bevölkerungswert quadriert werden.

Neben dem Ausmaß und der Richtung der Prognosefehler stellt auch die (Un)Ähnlichkeit des Verlaufs der realen und der prognostizierten Bevölkerungsentwicklung ein Bewertungskriterium dar. Je ähnlicher sich die beiden Verläufe sind, desto höher ist die Prognosekraft zu bewerten. Eine Möglichkeit zur Bestimmung der Ähnlichkeit der Verläufe stellt eine lineare Regression dar, mit der die Abstände zwischen den realen und prognostizierten Bevölkerungswerten (z.B. mit den Konstanten  $c$  und  $\beta$  über die Zeit  $t$ ) modelliert werden können (vgl. Formel 8.16). Ist die Güte der linearen Regression<sup>7</sup> zu niedrig, unterscheiden sich die beiden Verläufe zu stark. In der Abbildung 8.1 ist dieser Sachverhalt noch einmal graphisch veranschaulicht.

$$A_t = F_{p_i,t} + c + \beta \cdot t \quad (8.16)$$

### Zusammenfassung

Fehlermaße zur Bewertung probabilistischer Bevölkerungsprognosen können überprüfen, inwiefern die realen Bevölkerungswerte prognostiziert worden sind. Dazu kann die Ergebnisverteilung hinsichtlich Ausmaß, Richtung, Streuung und Verlaufsstruktur der Prognosefehler untersucht werden. Allerdings können daraus weder Rückschlüsse auf die getroffenen

---

<sup>7</sup>Ein Gütemaß einer linearen Regression ist beispielsweise das Bestimmtheitsmaß.



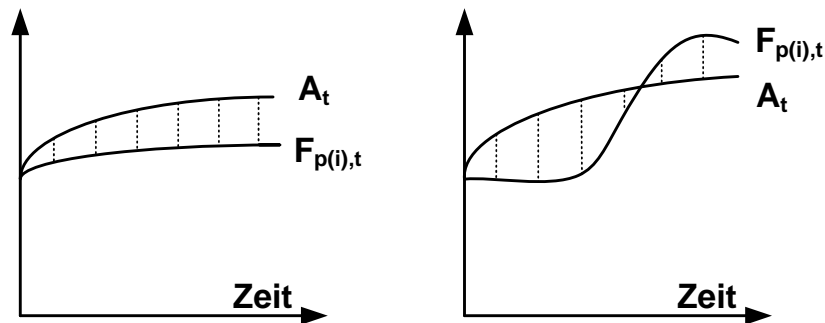


Abbildung 8.1.: (Nicht)Modellierbarkeit der Abweichungen ähnlicher (links) und unähnlicher (rechts) Bevölkerungsentwicklungen ( $A_t$  und  $F_{p(i),t}$ ) mit einer linearen Regression

Annahmen noch auf die Prognosemethode gezogen werden.<sup>8</sup>

Um eine Ergebnisverteilung probabilistischer Bevölkerungsprognosen im Querschnitt wie auch im Längsschnitt umfassend bewerten zu können, sollten mehrere der in diesem Abschnitt entwickelten Fehlermaße verwendet werden. So gibt der *LARD* die Position des realen Wertes innerhalb (wie auch außerhalb) der Ergebnisverteilung an. Außerdem können der *MALPERD* (oder der *MERD*) und der *MAPERD* (oder der *MAERD*) Richtung, Ausmaß und Streuung des Prognosefehlers in einem Projektionsintervall bestimmen. Werden diese Fehlermaße anstatt für ein Projektionsintervall über den gesamten Projektionshorizont betrachtet, so können zudem Erkenntnisse über die Entwicklung des Prognosefehlers im Verlauf hinsichtlich des Niveaus und der Verteilung analysiert werden. Daneben kann auch die Ähnlichkeit der realen und der prognostizierten Bevölkerungsentwicklung geprüft werden, indem die Abweichungen mit einer linearen Regression modelliert werden. Je höher die Güte der linearen Regression, desto ähnlicher sind sich die Verläufe und desto höher ist die Prognosekraft (des Modells).

## 8.2. Komplexe versus simple Prognosemodelle

In der Demographie gibt es im Bereich der Bevölkerungsprognosen eine lang anhaltende Diskussion darüber, ob simple Bevölkerungsprognosemodelle besser oder schlechter sind als

---

<sup>8</sup>Um Rückschlüsse auf die Annahmen ziehen zu können, müsste die *wahre* Wahrscheinlichkeitsverteilung der Ereignisse bekannt sein. Um Rückschlüsse auf die Prognosemethode ziehen zu können, müssten zahlreiche Bevölkerungsprognosen mit der Prognosemethode wiederholt ausgeführt und ausgewertet werden.

komplexe Bevölkerungsprognosemodelle.

Simple und komplexe Bevölkerungsprognosemodelle unterscheiden sich in wesentlichen Dimensionen wie z. B. der Anzahl und dem Aggregationsniveau der Modellparameter oder auch in der Wahl der Methode(n) zur Generierung der Annahmen (vgl. Abschnitt 4.2 ab Seite 57) [169]. Simple Bevölkerungsprognosemodelle nutzen häufig nur wenige Modellparameter auf einem hohen Aggregationsniveau, um die Bevölkerung oftmals nur anhand ihrer vergangenen Entwicklung in die Zukunft fortzuschreiben bzw. zu extrapolieren. Dagegen nutzen komplexe Bevölkerungsprognosemodelle mehrere Modellparameter auf einem niedrigen Aggregationsniveau, um die Bevölkerung unter Berücksichtigung vielfältiger Informationsquellen, die sich in einem Mix verschiedener Methoden zur Annahmengenerierung widerspiegeln können, zu prognostizieren.

Die entscheidende Frage bei der Diskussion über die bessere Eignung simpler oder komplexer Bevölkerungsprognosemodelle geht mit der Gegenüberstellung von Aufwand und Leistung einher. Insofern die Prognoseergebnisse simpler und komplexer Bevölkerungsprognosemodelle unter denselben Prognosebedingungen in der Genauigkeit wie auch im Nutzen für den Endverbraucher übereinstimmen, sind simple komplexen Bevölkerungsprognosemodellen vorzuziehen, da der Mehraufwand eines komplexen Bevölkerungsprognosemodells dann ungerechtfertigt ist. Wenn allerdings komplexe Bevölkerungsprognosemodelle genauere und für den Endverbraucher geeignetere Prognoseergebnisse erzeugen, so sind sie simplen Bevölkerungsprognosemodellen vorzuziehen.<sup>9</sup>

Es gibt bislang keinen eindeutigen (und reliablen) Test, um die jeweils geeignetsten Prognosebedingungen für simple und komplexe Bevölkerungsprognosemodelle zu identifizieren. Testergebnisse variieren z. B. in Abhängigkeit des zu verwendenden Fehlermaßes, der Basisperiode und der Länge des Projektionshorizonts. Außerdem haben Untersuchungen gezeigt, dass es bei solchen Tests schwierig ist, zu unterscheiden, ob die beobachtete höhere Genauigkeit simpler oder komplexer Bevölkerungsprognosemodelle auf die jeweils herrschenden (in)stabilen demographischen Verhältnisse oder tatsächlich auf die Prognosemethode zurückzuführen ist [238]. In diesem Zusammenhang stellt Willekens [307] fest,

---

<sup>9</sup>So stellt z. B. Long [169] fest, dass simple Bevölkerungsprognosemodelle für die Prognose aggregierter Ergebnisgrößen (wie z. B. der totalen Bevölkerungszahl) geeignet sind, wohingegen komplexe Bevölkerungsprognosemodelle für die Prognose disaggregierter Ergebnisgrößen (wie z. B. der totalen Bevölkerung nach Alter und Geschlecht) geeignet sind.

dass bei vergleichenden Tests simpler und komplexer Bevölkerungsprognosemodelle die (In)Stabilität der demographischen Bedingungen neben anderen Prognosebedingungen fälschlicherweise nachrangig behandelt wird. Simple Bevölkerungsprognosemodelle können unter stabilen demographischen Verhältnissen über einen kurzen Projektionshorizont noch ähnlich genaue Prognoseergebnisse erzielen wie komplexe Bevölkerungsprognosemodelle. Sobald sich jedoch die demographischen Verhältnisse verändern, können diese nur noch von komplexen Bevölkerungsprognosemodellen antizipiert und adäquat eingefangen werden. Dadurch wird der Nutzen komplexer Bevölkerungsprognosemodelle noch gesteigert, da insbesondere in instabilen Zeiten genaue Voraussagen zur Bevölkerungsentwicklung für diverse Endverbraucher vonnöten sind.

Simple (oder naive) Bevölkerungsprognosemodelle haben den Vorteil, relativ einfach umsetzbar zu sein. So benötigen sie relativ wenige Daten, sind einfach zu berechnen und können durch ihre Transparenz relativ gut überprüft werden. Allerdings sind simple Bevölkerungsprognosemodelle auch fehleranfällig, da sie lediglich ein Bruchstück der relevanten verfügbaren Informationen nutzen.<sup>10</sup> Komplexe Bevölkerungsprognosemodelle nutzen hingegen möglichst alle verfügbaren Informationsquellen, um disaggregierte Ergebnisgrößen genau zu prognostizieren. Dabei ist es unvermeidlich, die Modellstruktur entsprechend zu erweitern, um mehr Daten für mehr Modellparameter in einem anspruchsvolleren Rechenprozess nutzen zu können.

Demnach sind weder simple komplexen noch komplexe simplen Bevölkerungsprognosemodellen rigoros vorzuziehen. Vielmehr sollte sich die Wahl eines simplen oder komplexen Bevölkerungsprognosemodells an den jeweiligen Prognosebedingungen orientieren. Dazu zählen beispielsweise die eigentliche Fragestellung, die Anforderungen der Endverbraucher wie auch die Länge des Projektionshorizonts.

### 8.3. Schlussfolgerung — Anforderungen an das PPPM

Die Diskussion der Bewertbarkeit (und Bewertung) von Bevölkerungsprognosen wie auch die Gegenüberstellung komplexer und simpler Prognosemethoden hat zwei weitere interessante Aspekte (neben denen aus den Abschnitten 4.3, 5.3, 6.2 und 7.2) für die Entwicklung des

---

<sup>10</sup>So würde beispielsweise ein simples Bevölkerungsprognosemodell für zwei zwar gleichgroße, aber unterschiedlich alte Bevölkerungen dieselbe totale Bevölkerungszahl extrapolieren, obwohl die ältere Bevölkerung in der Zukunft wahrscheinlich kleiner sein würde als die jüngere Bevölkerung.

PPPM hervorgebracht.

So ist bei der Modellspezifikation des PPPM darauf zu achten, dass der Detailgrad der Modellparameter einerseits so hoch ist, dass auch bei instabilen Bedingungen noch eine genaue Prognose erstellt werden kann. Andererseits sollte der Detailgrad der Modellparameter aber auch so niedrig sein, dass mit den verfügbaren Daten (verschiedener Informationsquellen) immer noch plausible Annahmen generiert werden können. Dieser Aspekt betrifft beispielsweise die Modellierung der Migration. Häufig prognostizieren herkömmliche Bevölkerungsprognosemodelle vereinfachend die Nettomigration mit denselben Annahmen zur Fertilität und Mortalität für die Nettoimmigranten und die Einheimischen (vgl. Abschnitt 4.2 ab Seite 57). Im PPPM soll nun aber Bruttomigration mit separaten Annahmen zur Fertilität und Mortalität berücksichtigt werden, sodass die direkten Zu- und Fortzüge und ihre Nachkommengenerationen einzeln prognostiziert werden können. Dadurch kann beispielsweise ein sukzessiver Anpassungsprozess von der höheren Fertilität der direkten Zuzüge an das niedrigere Niveau der Einheimischen über die einzelnen Nachkommengenerationen simuliert werden (vgl. Abschnitt 7.2). Da sich nun aber das demographische Verhalten nicht nur zwischen, sondern auch in diesen Subpopulationen unterscheidet, wäre es noch realistischer, die Annahmen zur Fertilität und Mortalität zusätzlich in Abhängigkeit von dem jeweiligen Herkunfts- bzw. Zielland zu generieren. Allerdings wäre dann der Detailgrad bei den Modellparametern der Migration so hoch, dass kaum noch plausible Annahmen zu deren Fertilität und Mortalität aufgrund fehlender detaillierter Daten generiert werden können. Folglich wird im PPPM der Detailgrad der Modellparameter nicht nur in Abhängigkeit von theoretischen und empirischen Erkenntnissen, sondern auch in Abhängigkeit von den verfügbaren Daten bestimmt, sodass die Prognosegenauigkeit angemessen gesteigert werden kann.

Daneben tritt die Erkenntnis, dass die Prognosegüte eines Prognosemodells nicht ausschließlich über den Abgleich der prognostizierten und tatsächlich eingetroffenen Werte erfolgen kann, da die *wahre* Wahrscheinlichkeitsverteilung der potentiell in der Zukunft eintretenden Ereignisse weder vor noch nach einer Prognose bekannt ist. So kann z. B. eine Bevölkerungsprognose fälschlicherweise als ungenau eingeschätzt werden, wenn ein unwahrscheinlich prognostiziertes Ereignis tatsächlich eintritt, obwohl es auch in der *wahren* Wahrscheinlichkeitsverteilung nur eine geringe Eintrittswahrscheinlichkeit hat. Folglich sollte bei probabilistischen Bevölkerungsprognosemodellen neben den Prognoseergebnissen vor allem die Fähigkeit,

alle potentiellen zukünftigen Entwicklungen der Modellparameter flexibel berücksichtigen zu können, bewertet werden.

Teil II.

# Probabilistic Population Projection Model (PPPM)

## 9. Grundlegende Modellparameter des PPPM

Im ersten Teil der Arbeit sind u. a. bestehende Ansätze zur probabilistischen Bevölkerungsprognose (vgl. Kapitel 4) wie auch zur Prognose einzelner Modellparameter der Mortalität (vgl. Kapitel 5), Fertilität (vgl. Kapitel 6) und Migration (vgl. Kapitel 7) betrachtet und hinsichtlich ihrer Stärken und Schwächen analysiert worden. Diese Analyse bisheriger Prognoseansätze hat zahlreiche Anforderungen zur Entwicklung des neuartigen Prognosemodells PPPM hervorgebracht, um die zukünftige Bevölkerungsentwicklung umfassender und flexibler (als herkömmliche Bevölkerungsprognosemodelle) prognostizieren zu können.

In den folgenden Kapiteln 9 bis 11 werden nun das PPPM und dessen prototypische Realisierung als Software, das P3J, systematisch vorgestellt; von den Modellparametern über die Berechnung wesentlicher Ergebnisgrößen bis hin zur Analyse der Prognoseergebnisse (vgl. Abbildung 9.1).

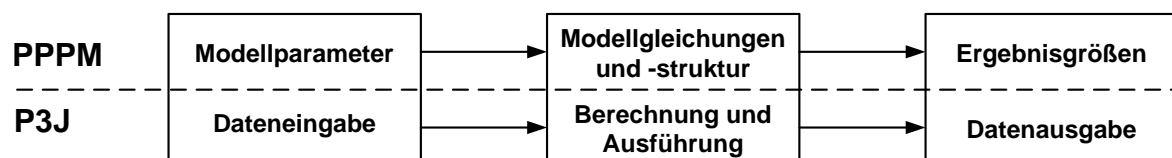


Abbildung 9.1.: Aspekte des theoretischen Modells, PPPM, und seiner Realisierung P3J

Jedes dieser drei Kapitel ist ähnlich aufgebaut: Anfangs wird der konzeptionelle Aufbau des theoretischen Modells — PPPM — demographisch motiviert und mathematisch formuliert. Anschließend wird dessen praktische Umsetzung im P3J erläutert.

Das P3J ist eine implementierte Realisierung des PPPM, die in enger Zusammenarbeit mit dem Lehrstuhl für Modellierung und Simulation an der Universität Rostock über mehrere

Jahre entstanden ist. Anforderungen an die zu realisierende Software sind sukzessive aus der im Vordergrund stehenden konzeptionellen Entwicklung des PPPM entstanden. Dabei sind u. a. Modellgleichungen, Organisationsstrukturen zur Kombination der Modellparameter und Annahmenverläufe sowie Algorithmen zur simulativen und (quasi-)vollständigen Berechnung zunächst theoretisch entwickelt und dann z. T. mit Hilfe des Lehrstuhls für Modellierung und Simulation in Java programmiert worden. Ähnlich verhält es sich mit der Entwicklung der Bedienoberfläche des P3J: Aufbau und Inhalt sind zunächst theoretisch (mit diversen Skizzen) entwickelt und anschließend mit Unterstützung umgesetzt worden. Die praktische Umsetzung hat jeweils aus einem mehrstufigen Austauschprozess bestanden, wobei mehrmalige intensive Testphasen immer wieder Hinweise für Verbesserungen geliefert haben. Programmbausteine, die auf R und  $\text{\LaTeX}$  beruhen, wie z. B. beim Ergebnisreport (vgl. Abschnitt 11.2 ab Seite 226), sind ohne unterstützende Hilfe entstanden.

In diesem Kapitel werden nun der Aufbau des PPPM nach Subpopulationen wie auch die sich daraus ergebenden Modellparameter demographisch und formal (mathematisch) erläutert. (Die hier aufgeführten Modellparameter sind allerdings nicht vollständig, da die Erläuterung weiterer Modellbausteine im folgenden Kapitel 10 die Einführung noch weiterer Modellparameter nach sich führt.) Abschließend wird die Eingabe der Daten grundlegender Modellparameter im P3J geschildert (vgl. Abbildung 9.2).

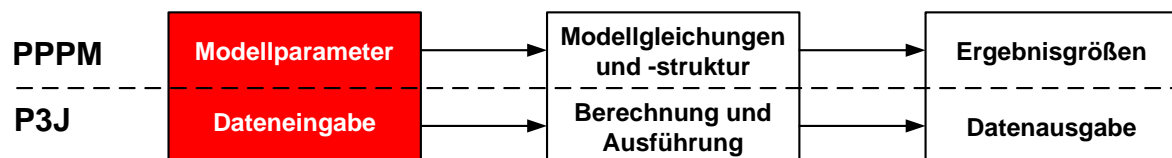


Abbildung 9.2.: Aspekte des theoretischen Modells, PPPM, und seiner Realisierung P3J: Das aktuelle Kapitel 9 befasst sich mit den Modellparametern des PPPM bzw. der Dateneingabe im P3J (rot unterlegt).

## 9.1. Zu unterscheidende Subpopulationen im PPPM

Im probabilistischen Bevölkerungsprognosemodell PPPM setzt sich die Gesamtbevölkerung aus diversen Subpopulationen zusammen. Zu den separat (nach Geschlecht und Einzelalter) zu prognostizierenden Subpopulation zählen die Einheimischen, die direkten Zuzüge und ihre Nachkommengenerationen wie auch die direkten Fortzüge und ihre Nachkommengene-



rationen (vgl. Abbildung 9.3).

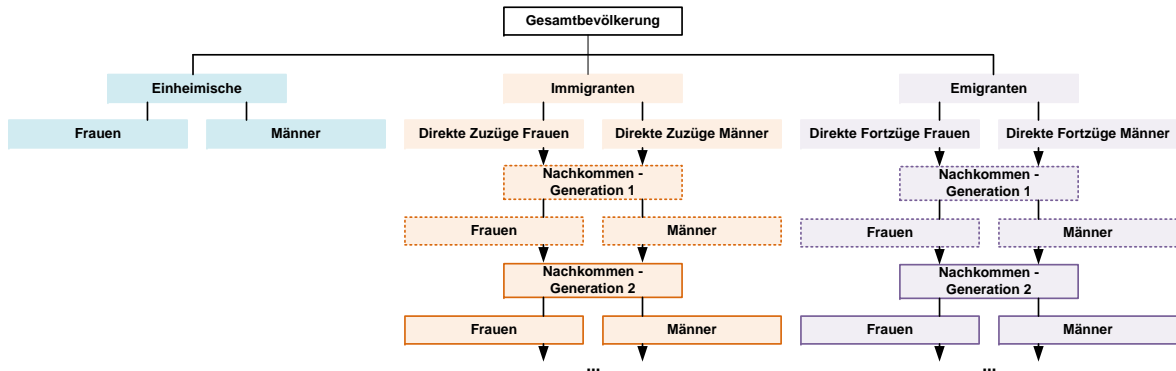


Abbildung 9.3.: Aufbau des PPPM hinsichtlich zu unterscheidender Subpopulationen: die Gesamtbevölkerung setzt sich aus den Einheimischen (blau), den Immigranten (orange) und Emigranten (lila) zusammen.

Jede Subpopulation wird mit separaten Annahmen zur Fertilität und Mortalität einzeln prognostiziert. Folglich kann beispielsweise im Gegensatz zu herkömmlichen Bevölkerungsprognosemodellen (vgl. Abschnitt 4.2 ab Seite 57) ein Anpassungsprozess von der höheren Fertilität oder/und niedrigeren Mortalität der Immigranten an das jeweilige Niveau der Einheimischen über die einzelnen Nachkommengenerationen modelliert werden (vgl. Abschnitt 7.1.2 ab Seite 138).

Darüber hinaus kann die Anzahl der bei einer Bevölkerungsprognose mit dem PPPM zu unterscheidenden Nachkommengenerationen der Zu- und Fortzüge in Abhängigkeit von der Länge des Projektionshorizonts (und dem Generationenabstand) variabel bestimmt werden.

Nachfolgend soll kurz erläutert werden, welche Personen zu welcher Subpopulation gehören.

### 9.1.1. Jump-off Bevölkerung

Die *Jump-off Bevölkerung* stellt die Gesamtbevölkerung zum Ende des Jahres dar, von dem aus die Bevölkerungsprognose startet. Generell entspricht sie der *de jure* Bevölkerung in einer genau definierten Gebietseinheit zu einem bestimmten Stichtag.<sup>1</sup>

<sup>1</sup>Da die Erfassung des Bevölkerungsbestandes nach dem *de jure* Konzept international uneinheitlich ist [83], sollten die entsprechenden länderspezifischen Praktiken bei einer Bevölkerungsprognose angemessen

Bei einer aktuellen Bevölkerungsprognose in die Zukunft kommen die Daten für die Jump-off Bevölkerung für gewöhnlich aus dem Jahr mit den letzten verfügbaren Daten.

### 9.1.2. Einheimische

Ausgehend von der Jump-off Bevölkerung entwickelt sich die Subpopulation der *Einheimischen* ausschließlich über in ihr stattfindende Geburten und Sterbefälle. Demzufolge kann die Bevölkerung der Einheimischen weder über direkte Zu- bzw. Fortzüge noch über deren Nachkommen im In- bzw. Ausland wachsen bzw. schrumpfen. Anstatt dessen ist allein die natürliche Bevölkerungsbewegung in der einheimischen Bevölkerung für ihre numerische und bevölkerungsstrukturelle Veränderung ausschlaggebend.

### 9.1.3. Immigranten

Zu den Immigranten zählen auf einer ersten Ebene die *direkten Zuzüge*. Unter direkten Zuzügen werden (neben (in der Gebietseinheit gebürtigen) Remigranten) Personen zusammengefasst, die im Ausland geboren worden sind und die in die zu betrachtende, genau definierte Gebietseinheit ziehen. Die Subpopulation der direkten Zuzüge verändert sich in ihrem Bestand und in ihrer Bevölkerungsstruktur durch Annahmen zu den stattfindenden direkten Zuzügen wie auch zur Sterblichkeit der direkten Zuzüge im Inland über den Projektionshorizont.

Die im Inland von den direkten Zuzügen zu realisierenden Geburten bilden die erste Nachkommengeneration der direkten Zuzüge, wohingegen die Kinder der ersten Nachkommengeneration bereits die zweite Nachkommengeneration der ursprünglich direkt Zugezogenen sind. In der Folge können in Abhängigkeit von der Länge des Projektionshorizonts und des anzunehmenden Generationenabstands über die Fertilität  $n$  Nachkommengenerationen der direkten Zuzüge in einer Bevölkerungsprognose mit dem PPPM berücksichtigt werden. Bestand und Struktur einer Nachkommengeneration  $n$  der direkten Zuzüge verändern sich durch die Geburten der vorangegangenen Nachkommengeneration  $n - 1$  bzw. der direkten Zuzüge und durch die anzunehmende Sterblichkeit über den Projektionshorizont.

---

berücksichtigt werden.

#### 9.1.4. Emigranten

Zu den Emigranten zählen auf einer ersten Ebene die *direkten Fortzüge*. Unter direkten Fortzügen werden Personen zusammengefasst, die im In(- oder Aus-)land geboren worden sind und vom Inland ins Ausland ziehen. Die Subpopulation der direkten Fortzüge verändert sich in ihrem Bestand und in ihrer Bevölkerungsstruktur durch Annahmen zu den stattfindenden direkten Fortzügen wie auch zur Sterblichkeit der direkten Fortzüge über den Projektionshorizont.

Die im Ausland von den direkten Fortzügen zu realisierenden Geburten bilden die erste Nachkommengeneration der direkten Fortzüge, wohingegen die Kinder der ersten Nachkommengeneration bereits die zweite Nachkommengeneration der ursprünglich direkt Fortgezogenen sind. In der Folge können in Abhängigkeit von der Länge des Projektionshorizonts und des anzunehmenden Generationenabstands über die Fertilität  $n$  Nachkommengenerationen der direkten Fortzüge in einer Bevölkerungsprognose mit dem PPPM berücksichtigt werden. Bestand und Struktur einer Nachkommengeneration  $n$  der direkten Fortzüge verändern sich durch die Geburten der vorangegangenen Nachkommengeneration  $n - 1$  bzw. der direkten Fortzüge und durch die anzunehmende Sterblichkeit über den Projektionshorizont.

#### 9.1.5. Gesamtbevölkerung

Zur Bildung der Gesamtbevölkerung in einem Projektionsintervall werden die direkten Fortzüge und ihre Nachkommengenerationen von den Einheimischen und den direkten Zuzügen und ihren Nachkommengenerationen abgezogen.

Die Prognose der einzelnen Subpopulationen ermöglicht zwar separate Fertilitäts- und Mortalitätsannahmen sowie die Analyse vielfältiger bevölkerungsdynamischer Effekte auf einzelne oder zusammengeschlossene Subpopulationen (vgl. Kapitel 11 und Abschnitt 12.4 ab Seite 242). Allerdings kann sie auch bei der Zusammenfassung der Subpopulationen zur Gesamtbevölkerung Unstimmigkeiten hervorrufen. So werden die direkten Fortzüge und ihre Nachkommengenerationen durch die separate Prognose der Subpopulationen nicht sofort im Laufe der Prognose, sondern erst am Projektionsende bei den Einheimischen und den Immigranten abgezogen. Da allerdings alle Subpopulationen eigenen, anzunehmenden Fertilitäts- und Mortalitätsbedingungen über den Projektionshorizont ausgesetzt sind, können die separat prognostizierten direkten Fortzüge und ihre Nachkommengenerationen von den in den ande-

ren Subpopulationen mitgeführten direkten Fortzügen und ihren Nachkommengenerationen abweichen. Bei den direkten Zuzügen und ihren Nachkommengenerationen besteht diese Diskrepanz nicht, da sie bei den Einheimischen im Laufe der Prognose nicht mitgeführt werden können, sondern einfach fehlen. Demzufolge müssen am Ende der Prognose bei der Konstruktion der Gesamtbevölkerung keine separat prognostizierten und mitgeführten Bestände der direkten Zuzüge und ihrer Nachkommengenerationen gegeneinander aufgerechnet werden. Die hier beschriebene Diskrepanz bei den Emigranten ist zwar theoretisch von Belang, spielt aber numerisch eine untergeordnete Rolle; generell gilt, je stärker sich die Annahmen zur Fertilität und Mortalität zwischen den Emigranten, Einheimischen und Immigranten unterscheiden, desto größer wird diese Diskrepanz sein.

#### 9.1.6. Theoretischer Hintergrund

Die Idee, eine zu prognostizierende Bevölkerung in relevante Subpopulationen zu zerlegen und separat zu prognostizieren, ist nicht neu. So beschäftigen sich u. a. Espenshade et al. [82, 15], Mitra [194, 196], Schmertmann [247], Cerone [55] und Edmonston und Passel [79] mit den Auswirkungen von Zuwanderung unter besonderen bis hin zu allgemeineren Annahmen auf die zu prognostizierende Bevölkerung. Das Bevölkerungsprognosemodell von Edmonston und Passel [79] aus dem Jahr 1992 kommt dabei dem PPPM relativ nah, da es die Immigranten in einzelne Subpopulationen zerlegt. Demnach werden im Modell von Edmonston und Passel [79] die direkten Zuzüge von drei ihrer weiteren Nachkommengenerationen<sup>2</sup> unterschieden. Für jede dieser Subpopulationen können eigene Annahmen zur Fertilität, Mortalität und Migration getroffen werden. Allerdings werden beim Bevölkerungsprognosemodell von Edmonston und Passel [79] nur drei Nachkommengenerationen der direkten Zuzüge unterschieden, wohingegen beim PPPM die Anzahl der zu betrachtenden Nachkommengenerationen frei bestimmt werden kann und demnach keinen Beschränkungen unterliegt. Darüber hinaus werden beim PPPM im Gegensatz zum Modell von Edmonston und Passel [79] nicht nur die Immigranten, sondern auch die Emigranten detailliert mit direkten Fortzügen und entsprechenden Nachkommengenerationen modelliert und prognostiziert, wobei für jede dieser Subpopulationen separat Annahmen zur Mortalität und Fertilität getroffen werden können. Damit ermöglicht das PPPM vielfältige weiterführende Analysen bevölkerungsdynamischer Effekte diverser Zu- und Fortzugsszenarien als ein herausstechendes Merkmal.

---

<sup>2</sup>Die erste Nachkommengeneration sind die Kinder der direkten Zuzüge, die zweite Nachkommengeneration sind die Enkelkinder der direkten Zuzüge und die dritte Nachkommengeneration umfasst alle weiteren Kinder, die zur dritten oder höheren Nachkommengeneration gehören.

## 9.2. Formale Definition der zu verwendenden Modellparameter im PPPM

Bei einer Bevölkerungsprognose mit dem PPPM müssen zunächst in Abhängigkeit von der Fragestellung (vgl. Abbildung 2.2 auf Seite 19)

- die Länge des Projektionshorizonts in Jahren
- die Anzahl der zu unterscheidenden Einzelalter (wie auch)
- die Anzahl der zu unterscheidenden Nachkommengenerationen der Migranten

bestimmt werden.

Anschließend muss eine *Jump-off* Bevölkerung festgelegt werden, die den eigentlichen Ausgangspunkt einer Bevölkerungsprognose darstellt. Daraufhin müssen für jede Subpopulation Annahmen zu verschiedenen Modellparametern der Fertilität, Mortalität und Migration getroffen werden.

### 9.2.1. Jump-off Bevölkerung

Im PPPM muss eine *Jump-off* Bevölkerung ausschließlich für die einheimische Bevölkerung eingegeben werden. Dazu wird die Jahresendbevölkerung der zu prognostizierenden Wohnbevölkerung nach Alter und Geschlecht, meist für das Jahr mit den aktuellsten verfügbaren Daten, herangezogen.

### 9.2.2. Direkte Zu- und Fortzüge

Im PPPM müssen für die Modellparameter *direkte Zuzüge* und *direkte Fortzüge* absolute Bestandszahlen nach Alter und Geschlecht als Annahmenverläufe über den gesamten Projektionshorizont getroffen und in das Modell eingegeben werden. Dabei werden nur die Zu- und Fortzüge betrachtet, die direkt in die zu prognostizierende einheimische Bevölkerung einwandern bzw. die direkt aus der zu prognostizierenden einheimischen Bevölkerung auswandern.

### 9.2.3. Altersspezifische Fertilitätsraten

Für jede weibliche Subpopulation (der Einheimischen, Immi- und Emigranten) müssen Annahmen zu den *altersspezifischen Fertilitätsraten* über den gesamten Projektionshorizont

generiert und in das Modell eingegeben werden.

*Altersspezifische Fertilitätsraten* bestimmen die Anzahl der Geburten einer Frau in einem bestimmten Alter. Daher werden bei der Berechnung der *altersspezifischen Fertilitätsraten*,  $f_x$ , die Geburten,  $B$ , von Frauen in einem bestimmten Alter  $x$  zu der durchschnittlichen Anzahl der Frauen,  $\bar{P}^f$ , im Alter  $x$  in Beziehung gesetzt (vgl. Formel 9.1).

$$f_x = \frac{B_x}{\bar{P}_x^f} \quad (9.1)$$

#### 9.2.4. Überlebende im Alter x

Für jede Subpopulation müssen über den gesamten Projektionshorizont Annahmen zu den *Überlebenden im Alter x* generiert und in das Modell eingegeben werden. *Überlebende im Alter x* sind ein Sterbetafelparameter,<sup>3</sup> der die Anzahl der Personen einer fiktiven Sterbetafelbevölkerung im jeweils exakten Alter  $x, 0$  wiedergibt. Für das Alter 0 wird mit dem Radix die Größe der Sterbetafelbevölkerung zu Beginn festgelegt<sup>4</sup> [221]. Die Anzahl der Sterbetafelsterbefälle in den einzelnen Altersstufen richtet sich nach der tatsächlich beobachteten Sterblichkeit in der Allgemeinbevölkerung.

So werden die Überlebenden im exakten Alter  $x$ ,  $l_x$ , mit dem Produkt aus den Überlebenden im vorherigen Alter,  $l_{x-1}$ , und der Wahrscheinlichkeit, vom Alter  $x - 1$  bis zum Alter  $x$  zu überleben,  $1 - q_{x-1}$ , berechnet (vgl. Formel 9.2).

$$l_x = l_{x-1} \cdot (1 - q_{x-1}) \quad (9.2)$$

Die Sterbewahrscheinlichkeit  $q_x$  wird über den Quotienten der Sterbefälle  $D$  im Altersintervall  $[x, x + 1)$  und den Personen im exakten Alter  $x$ , die dem Sterberisiko am Beginn des Altersintervalls ausgesetzt sind, berechnet (vgl. Formel 9.3).

$$q_x = \frac{D_x}{P_x} \quad (9.3)$$

---

<sup>3</sup>Die Konstruktion einer Sterbetafel wird beispielsweise bei Preston et al. [221] ausführlich beschrieben.

<sup>4</sup>Der Radix kann willkürlich bestimmt werden. Üblich ist ein Radix von 100.000 Personen.

### 9.2.5. Geschlechteranteil bei Geburt

Annahmen zur Entwicklung des *Geschlechteranteils bei Geburt* müssen für alle Subpopulationen über den gesamten Projektionshorizont getroffen und in das Modell eingegeben werden. Der weibliche Geschlechteranteil bei Geburt,  $\pi(P_0^F)$ , bestimmt den Anteil der weiblichen Lebendgeborenen,  $LG^F$ , an allen Lebendgeborenen (vgl. Gleichung 9.4), wohingegen der männliche Geschlechteranteil bei Geburt,  $\pi(P_0^M)$ , den Anteil der männlichen Lebendgeborenen,  $LG^M$ , an allen Lebendgeborenen bestimmt (vgl. Formel 9.5).

$$\pi(P_0^F) = \frac{LG^F}{LG^F + LG^M} \quad (9.4)$$

$$\pi(P_0^M) = \frac{LG^M}{LG^F + LG^M} \quad (9.5)$$

### 9.2.6. Anteil der Säuglingssterbefälle im ersten Halbjahr

Für jede Subpopulation müssen Annahmen zum *Anteil der Säuglingssterbefälle im ersten Halbjahr* über den gesamten Projektionshorizont bestimmt und in das Modell eingegeben werden.

Der *Anteil der Säuglingssterbefälle im ersten Halbjahr*,  $\pi_{(0,5}D_0)$ , gibt den Anteil der Säuglinge, die im ersten halben Lebensjahr versterben,  $_{0,5}D_0$ , an allen im ersten Lebensjahr versterbenden Säuglingen,  $_1D_0$ , an (vgl. Formel 9.6).

$$\pi_{(0,5}D_0) = \frac{_{0,5}D_0}{_1D_0} \quad (9.6)$$

### 9.2.7. Überlebenswahrscheinlichkeit der obersten, nach oben offenen Altersgruppe

Im PPPM wird die Anzahl der Personen in einer frei wählbaren, obersten, nach oben offenen Altersgruppe mit dem Modellparameter *Überlebenswahrscheinlichkeit der obersten, nach oben offenen Altersgruppe* berechnet. Dieser Modellparameter gibt die Wahrscheinlichkeit an, mit der Personen dieser obersten Altersgruppe noch ein weiteres Projektionsintervall überleben. Die entsprechenden Annahmenverläufe sind extern und separat für jede Subpopulation über den Projektionshorizont zu generieren und einzugeben. Kannisto ([277], S. 33) hat in einer umfassenden Analyse der Sterblichkeit in den höchsten Altern kaum einen Wert für die Sterbewahrscheinlichkeit,  $q_x$ , gefunden, der 0,6 überschreitet. Weiterhin hat Gampe [90, 300]

mit Daten aus der internationalen Datenbank zur Langlebigkeit<sup>5</sup> ermittelt, dass die Sterbewahrscheinlichkeit der *Supercentenarians*, also der Personen, die mindestens 110 Jahre alt sind, den Wert von 0,5 nicht überschreitet. Folglich können solche Werte als Richtwerte bei der Generierung der Annahmen für den Modellparameter *Überlebenswahrscheinlichkeit der obersten, nach oben offenen Altersgruppe* herangezogen werden. Nachfolgend soll beispielhaft die Altersgruppe der 95+-Jährigen als oberste, nach oben offene Altersgruppe zur Beschreibung der definitorischen Zusammenhänge der Modellparameter des PPPM (vgl. Abschnitt 10.1 ab Seite 185) verwendet werden.

### 9.2.8. Überblick

Die Abbildung 9.4 zeigt den Aufbau des PPPM hinsichtlich der zu unterscheidenden Subpopulationen und der jeweils dazugehörenden Modellparameter.

## 9.3. P3J — Dateneingabe

Das PPPM ist in der prototypischen Anwendung P3J umgesetzt worden, um das zugrundeliegende theoretische Modell in der Praxis tatsächlich anwenden (und testen) zu können (vgl. Kapitel 12).

Nachfolgend soll die Realisierung des PPPM im P3J schrittweise erläutert werden. Dazu wird zunächst motiviert, warum das P3J auf der Grundlage des Modellierungs- und Simulationsrahmenwerks JAMES II entwickelt worden ist. Anschließend werden dann die intern verwendeten Datenstrukturen wie auch die graphische Benutzeroberfläche, die den Nutzer bei der Dateneingabe unterstützen soll, beschrieben.

### 9.3.1. Das P3J und James II

Ein komplexes Prognosemodell wie das PPPM als Software zu realisieren ist mit einem relativ großen Aufwand verbunden. Ein Grund dafür liegt in der Vielzahl von Funktionen, wie z. B. Mechanismen zur Speicherung, Darstellung und Bearbeitung von Modellparametern, die für eine sinnvolle Benutzung des PPPM nötig sind. Während die konkrete Ausgestaltung dieser Funktionen PPPM-spezifisch ist und daher neu implementiert werden muss, sind die *prinzipiellen* Funktionen (wie z. B. die Speicherung von Modellen oder deren Bearbeitung)

---

<sup>5</sup>International Database on Longevity (IDL)



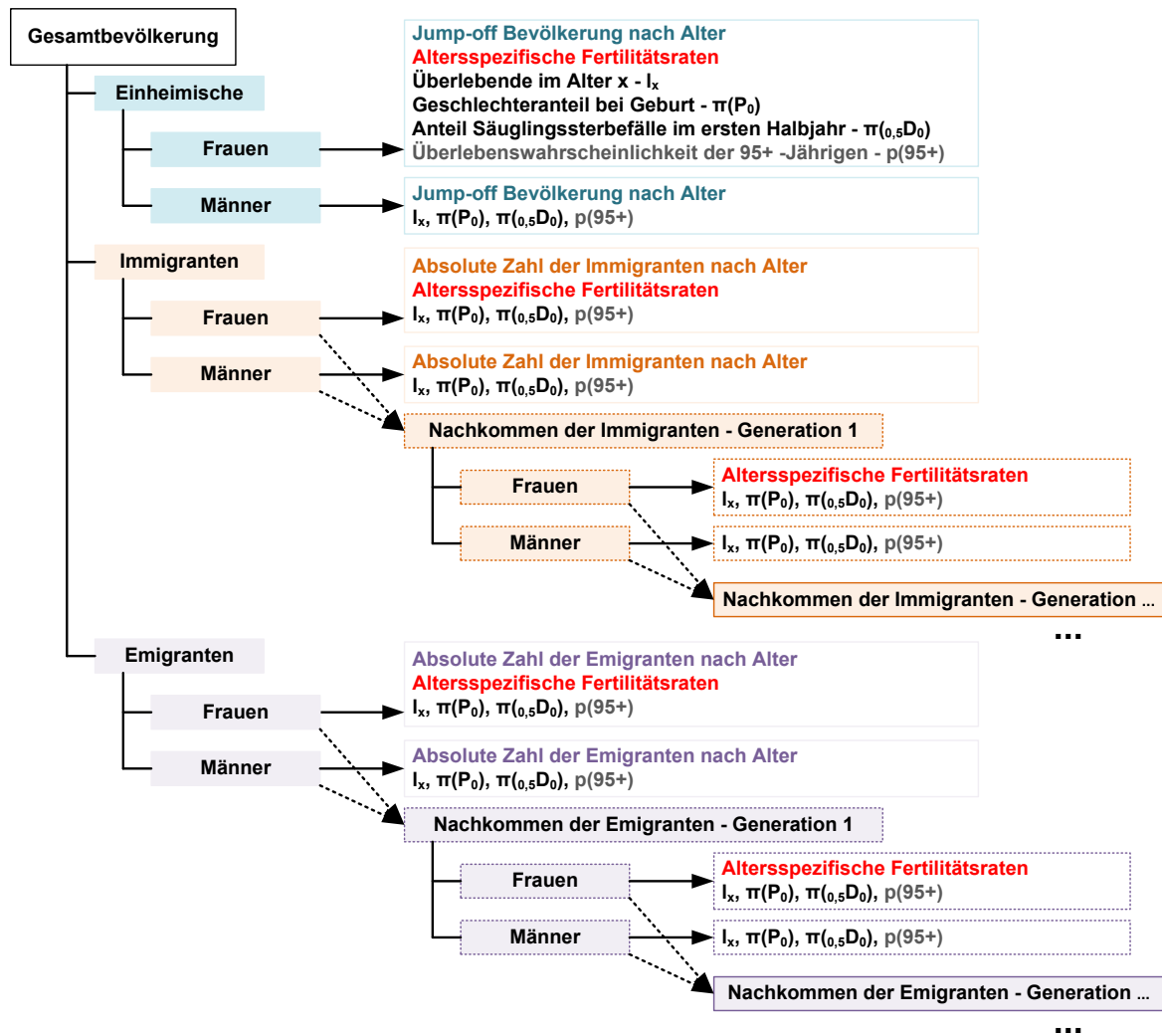


Abbildung 9.4.: Aufbau des PPPM hinsichtlich zu unterscheidender Subpopulationen und Modellparameter, wobei geschlechts- und subpopulationsspezifische Modellparameter extra farblich gekennzeichnet sind: Rot für die Frauen, hellblau für die Einheimischen, orange für die Immigranten und lila für die Emigranten; zudem ist der Modellparameter *Überlebenswahrscheinlichkeit der obersten, nach oben offenen Altersgruppe* mit der Farbe grau extra farblich gekennzeichnet, da er optional ist und die einzuschließenden Alter in dieser Altersgruppe selbst bestimmbar sind; beispielhaft sind in dieser Abbildung die 95+-Jährigen als oberste, nach oben offene Altersgruppe gewählt worden.

auch für viele andere Anwendungen wichtig.

Damit solche grundlegenden Funktionen nicht für jede Anwendung neu konzeptioniert werden müssen, werden Sammlungen von Hilfsfunktionen und vordefinierten Abläufen in so genannten Software-Rahmenwerken zusammengefasst. Im Gegensatz zu Software-Bibliotheken, die lediglich eine Menge von Hilfsfunktionen bereitstellen (, die dann explizit aufgerufen werden müssen), ermöglichen es die zusätzlich in einem Rahmenwerk definierten Arbeitsabläufe, auch den grundlegenden Kontrollfluss wiederzuverwenden, der vielen Funktionen eigen ist. Der Unterschied zwischen einem Rahmenwerk und einer Bibliothek wird z. B. beim Speichern von Dateien deutlich; während eine Bibliothek zwei Hilfsfunktionen zur Verfügung stellen könnte, wie z. B. eine zur Eingabe des gewünschten Dateinamens durch den Nutzer und eine weitere zum Speichern von Daten in einer Datei, könnte ein Rahmenwerk den Arbeitsablauf *Datei speichern* bereitstellen, der dann zunächst den gewünschten Dateinamen vom Nutzer abfragt und danach die Daten in die angegebene Datei schreibt.

Die Wiederverwendung von Funktionen und Abläufen kann die Entwicklungszeit von Anwendungssoftware drastisch verkürzen, d. h. auf Grundlage von Rahmenwerken können komplexe Anwendungen meist schneller realisiert werden. Neben der Zeitersparnis erhöht diese Herangehensweise auch die Verlässlichkeit der entwickelten Software, da die meisten Rahmenwerke separat (und in verschiedenen Kontexten) getestet werden. Der letztgenannte Aspekt ist für wissenschaftliche Software besonders relevant, da Programmierfehler hier leicht die gewonnenen Erkenntnisse verfälschen können [190].

Das *P3J* ist auf Basis des quelloffenen Modellierungs- und Simulationsrahmenwerks JAMES II [109] erstellt worden. JAMES II wird am Lehrstuhl für Modellierung und Simulation der Universität Rostock entwickelt. Es ist in der weit verbreiteten Programmiersprache Java [207] geschrieben. Dies erlaubt die zusätzliche Einbindung von weiteren Bibliotheken und Rahmenwerken, die in Java entwickelt worden sind. Außerdem können Java-Anwendungen auf den meisten Betriebssystemen ohne weitere Anpassungen ausgeführt werden, sodass das *P3J* relativ einfach installiert und benutzt werden kann. Das JAMES II für die Entwicklung von demographischen Simulationswerkzeugen geeignet ist, dokumentiert z. B. auch die Entwicklung von *MicMac* bzw. Mic-Core, welches ebenfalls von dessen Vorteilen Gebrauch macht [323] (vgl. Abschnitt 3.2.3 ab Seite 45).

Basis vom Rahmenwerk JAMES II ist das *Plug'n Simulate*-Konzept [110], welches es erlaubt, die in den Arbeitsabläufen verwendeten Komponenten — so genannte Plug-ins — auszutauschen. Durch das Hinzufügen von neuen, anwendungsspezifischen Plug-ins kann JAMES II daher schnell um die gewünschten Funktionen erweitert werden.

### 9.3.2. Interne Datenverwaltung

Die zahlreichen Modellparameter, die im PPPM definiert werden (siehe Abbildung 9.4), führen zu einer relativ großen Menge an Eingabedaten. Die genaue Anzahl der Parameter ist dabei von der Anzahl der zu betrachtenden Nachkommengenerationen (der direkten Zu- und Fortzüge) abhängig, sodass sie nicht von vornherein feststeht. Aus diesem Grund bietet es sich an, die verschiedenen *Arten* von Parametern, z. B. die altersspezifischen Fertilitätsraten (vgl. Abschnitt 9.2.3), zu unterscheiden. Für jede vorhandene Subpopulation wird automatisch ein entsprechender Modellparameter angelegt, der die altersspezifischen Fertilitätsraten für diese Subpopulation repräsentiert.

Die Annahmenverläufe, die mit einem bestimmten Modellparameter assoziiert werden, können als  $m \times n$  Matrizen aufgefasst werden.<sup>6</sup> Die Größe der Matrix, d. h. die Zeilenanzahl  $m$  und die Spaltenanzahl  $n$ , sowie die Bedeutung der darin enthaltenen Zahlen hängt vom konkreten Modellparameter ab. So werden die altersspezifischen Fertilitätsraten in einer Matrix gespeichert, bei der die Spalten die einzelnen Jahre des Projektionshorizonts und die Zeilen die verschiedenen Altersgruppen symbolisieren. Zusätzlich werden für jeden Annahmenverlauf noch weitere für den Nutzer hilfreiche Daten gespeichert; dazu zählen beispielsweise ein Name und ein Beschreibungstext. Diese Daten sollen den Nutzern helfen, sich schnell einen Überblick über die verwendeten Annahmenverläufe zu verschaffen. Um effektiv mit solchen Matrizen zu rechnen, werden die Matrix-Komponenten der frei verfügbaren COLT-Bibliothek [54] verwendet.

Zur Verwaltung großer Datenmengen eignen sich standardisierte Werkzeuge bzw. so genannte Datenbankmanagementsysteme (DBMS). Allerdings wird die Standardabfragesprache für Datenbankmanagementsysteme, die Structured Query Language (SQL), von vielen DBMS nur mit leichten Abwandlungen unterstützt. Das führt dazu, dass eine für ein bestimmtes DBMS geschriebene Anwendung nicht ohne Weiteres mit anderen DBMS zusammenarbeiten kann. Um dieses Problem zu vermeiden, verwendet das P3J die Java-

---

<sup>6</sup>Dies schließt auch Vektoren, d. h. Matrizen der Größe  $1 \times n$ , mit ein.

Persistenzschnittstelle HIBERNATE [108], die eine Vermittlungsposition zwischen DBMS und Anwendungsprogramm (wie hier dem P3J) einnimmt. Damit ist es möglich, das P3J mit diversen DBMS zu kombinieren und es somit an die jeweiligen Speicher-Anforderungen des Nutzers anzupassen.

Für kleinere Datenmengen, z. B. für eine einzelne Prognose mit wenigen Annahmenverläufen pro Modellparameter, reicht ein einfaches, lokal installiertes DBMS aus, welches die zu speichernden Daten z. B. in einfache Textdateien schreibt. Für umfangreichere Prognosen (mit relativ vielen Annahmenverläufen pro Modellparameter), die zudem in Gruppenarbeit erstellt werden, bietet sich hingegen die Verwendung eines zentral (über ein Netzwerk) verfügbaren und leistungsfähigen DBMS an. Neben der erhöhten Flexibilität für den Nutzer sichert die Verwendung eines DBMS auch die Übertragbarkeit der Daten in andere Softwaresysteme. Das P3J ist bisher mit dem quelloffenen DBMS MySQL [209, 231] genutzt und getestet worden.

### 9.3.3. Nutzeroberfläche zur Eingabe der Daten

Neben der Speicherung der Eingabedaten stellt auch die Gestaltung einer graphischen Benutzeroberfläche eine zentrale Herausforderung dar. Sie sollte darauf ausgerichtet sein, dem Nutzer die Dateneingabe so einfach wie möglich zu machen und außerdem einen guten Überblick über die bereits eingegebenen Daten erlauben. Dies schließt auch den *Inhalt* der Annahmenverläufe mit ein. Durch die Neuartigkeit der vorgestellten Prognosetechnik scheint es außerdem hilfreich, Nutzer in die Lage zu versetzen, Annahmenverläufe aus vorherigen PPPM-Prognosen abzufragen (und gegebenenfalls wiederzuverwenden).

Die Bedienoberfläche des P3J ist so gestaltet, dass der Nutzer mit wenigen Mausklicks einen gewünschten Modellparameter auswählen und für diesen Parameter Annahmenverläufe eingeben oder bearbeiten kann. Annahmenverläufe und Modellparameter werden, wie bei einem Dateisystem, in einer Baumstruktur angeordnet. Die unterste Ebene dieser Hierarchie bilden die Annahmenverläufe selbst, die mit jeweils einem Modellparameter assoziiert sind. Modellparameter stellen die nächste Hierarchieebene dar. Sie sind wiederum der Art der Subpopulation (z. B. Einheimische, Emigranten) und der betreffenden Nachkommengeneration zugeordnet (sofern dies relevant ist), wodurch sich eine weitere Hierarchieebene definieren lässt. Analog zum Windows Explorer (, dem Dateimanager von Microsoft Windows), der den meisten Anwendern geläufig sein sollte, wird die so definierte Baumstruktur in einem

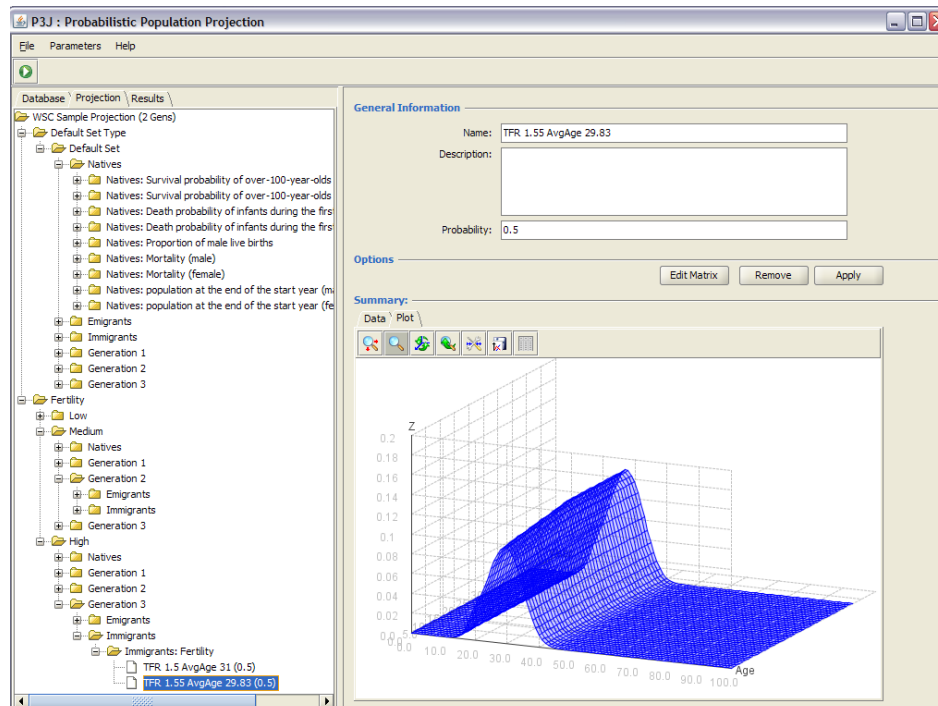


Abbildung 9.5.: Screenshot der P3J Eingabemaske für einen Annahmenverlauf: Der Nutzer wählt einen Modellparameter aus der Baumstruktur auf der linken Seite, um ihn dann im zentralen Bereich der Nutzeroberfläche zu bearbeiten.

abgetrennten Bereich auf der linken Seite angezeigt, während im Zentrum der *Inhalt* des jeweils links ausgewählten Elements angezeigt wird. Diese grundlegende Aufteilung zeigt der Screenshot in Abbildung 9.5.

Der Nutzer kann die Hierarchieebenen auf- und zuklappen, indem die entsprechenden Symbole im Baum auf der linken Seite ausgewählt werden. Wird ein Element der Hierarchie ausgewählt, werden seine Eigenschaften im Zentrum der Oberfläche angezeigt und können vom Nutzer verändert werden. Abbildung 9.5 zeigt dies für einen einzelnen Annahmenverlauf. Hat ein Element der Hierarchie untergeordnete Elemente, werden diese ebenfalls im Zentrum angezeigt und können vom Nutzer mit der Maus zur genaueren Ansicht ausgewählt werden. Diese in Analogie mit Webbrowsern gestaltete Funktion ermöglicht es Nutzern, sich schnell und einfach durch die komplexen Eingabeparameter des PPPM zu bewegen — ohne dabei den Überblick zu verlieren. Die Hierarchie auf der linken Seite zeigt immer das aktuelle Element (und seine Position in der Gesamtstruktur) an.

Die Nutzeroberfläche basiert auf Java Swing-Komponenten [208] aus der offiziellen Standard-Bibliothek für graphische Bedienelemente. Darüber hinaus verwendet die Nutzeroberfläche einige frei verfügbare Hilfsbibliotheken. Die Darstellung des in der Abbildung 9.5 ausgewählten Annahmenverlaufs erfolgt graphisch in einem Plot mit JMATHTOOLS [232] und tabellarisch mit JEPPERS [322]. Die Handhabung der tabellarischen Darstellung ist dabei stark an die Funktionalität von Tabellenkalkulationsprogrammen (wie Calc (OpenOffice.org) oder Microsoft Excel) angelehnt, z. B. hinsichtlich der Tastenkombinationen zum Selektieren und Einfügen von Daten. Dies soll die Benutzung der Oberfläche zusätzlich erleichtern.

## 10. Berechnung und Ausführung des PPPM

In diesem Kapitel werden sowohl die (deterministischen) Modellgleichungen (zur Berechnung grundlegender Ergebnisgrößen) als auch alternative Modi z. B. zur Erstellung und Kombination der Annahmenverläufe eines oder/und mehrerer Modellparameter wie auch zur simulativen oder/und (quasi-)vollständigen Berechnung des PPPM und P3J erläutert (vgl. Abbildung 10.1).

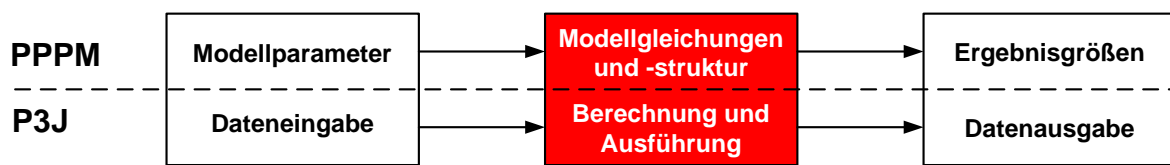


Abbildung 10.1.: Aspekte des theoretischen Modells, PPPM, und seiner Realisierung P3J:  
Das aktuelle Kapitel 10 befasst sich mit der Berechnung grundlegender Ergebnisgrößen des PPPM und P3J (rot unterlegt).

Dieses Kapitel ist dabei so aufgebaut, dass zunächst die deterministischen Modellgleichungen zur Berechnung wesentlicher Ergebnisgrößen im Abschnitt 10.1 vorgestellt werden; in einem deterministischen Modell wird ein Annahmenverlauf pro Modellparameter eingegeben, sodass auch nur ein Ergebnisverlauf pro Ergebnisgröße resultiert (vgl. Abschnitt 3.2.2 ab Seite 41).

Um nun anstatt eines Ergebnisverlaufs eine Ergebnisverteilung zu erhalten, wird in Abschnitt 10.2 erläutert, wie im PPPM (auf einzigartige Weise) mehrere Annahmenverläufe pro Modellparameter eingegeben, flexibel gewichtet und letztlich in mehreren Prognosedurchläufen kombiniert werden können.

Im Abschnitt 10.2.1 wird deshalb gezeigt, wie eine Annahmenverteilung erstellt werden

kann, die alle potentiellen Entwicklungen beinhalten und flexibel — entsprechend den Erwartungen eines Prognostikers — gewichten kann. Um diese Anforderungen zu erfüllen, werden im PPPM (im Gegensatz zu herkömmlichen probabilistischen Ansätzen) weder modell-basierte Methoden noch Verteilungen zur Generierung der Annahmenverläufe und Eintrittswahrscheinlichkeiten fest vorgegeben.

Darüber hinaus werden im Abschnitt 10.2.2 der *Open Type* und der *Limited Type* des PPPM vorgestellt, die sich in ihrer Fähigkeit unterscheiden, Annahmenverläufe eines oder/und mehrerer Modellparameter miteinander zu kombinieren. Im so genannten *Open Type* können alle Annahmenverläufe aller Modellparameter uneingeschränkt miteinander kombiniert werden, sodass auch unplausible Annahmenverlaufskombinationen, die nicht (theorie)konsistent sind oder den Erwartungen eines Prognostikers entsprechen, entstehen können. Um nun diese unplausiblen Kombinationen gezielt zu eliminieren, werden im so genannten *Limited Type* Settypen und Sets als Ordnungsstrukturen neu eingeführt, mit denen ein Prognostiker entsprechend seiner Erwartungen Annahmenverläufe eines oder mehrerer Modellparameter zu ausschließlich plausiblen Kombinationen zusammenschließen kann.

Im Abschnitt 10.3.2 werden dann noch verschiedene Modi zur Berechnung einer Ergebnisverteilung vorgestellt, die sowohl in Abhängigkeit von der Anzahl der Annahmenverläufe als auch von der eigentlichen Zielstellung optional gewählt werden können. Bei heterogenen Annahmenverteilungen mit relativ vielen Annahmenverläufen pro Modellparameter ist die simulative Berechnung zu wählen, wohingegen die vollständige Berechnung bei homogenen Annahmenverteilungen mit relativ wenigen Annahmenverläufen pro Modellparameter zu wählen ist. Die (quasi-)vollständige Berechnung ist hingegen bei der erweiterten Ergebnisanalyse vorteilhaft; insbesondere wenn anstatt der gesamten Ergebnisverteilung nur die  $n$  wahrscheinlichsten Ergebnisverläufe ausgegeben werden sollen.

Abschließend wird im Abschnitt 10.4 die Umsetzung dieser konzeptionellen Bausteine des PPPM zur Erzeugung einer Ergebnisverteilung für jede Ergebnisgröße im P3J erklärt.



### 10.1. Berechnung *einer* Annahmenverlaufskombination bzw. eines *Ergebnisverlaufs*

Im PPPM werden die Bestände der Subpopulationen zwischen dem Start- und Zieljahr in Halbjahresschritten prognostiziert. Folglich stellen Jahresend- und Jahresdurchschnittsbestände wesentliche Ergebnisgrößen der einzelnen Subpopulationen dar.

Generell finden demographische Ereignisse wie Sterbefälle, Geburten, Zuzüge und Fortzüge über das ganze Jahr verteilt statt. Allerdings muss bei der Berechnung der wesentlichen Ergebnisgrößen im PPPM eine Reihenfolge der demographischen Ereignisse festgelegt werden. Um dabei die Kontinuität der demographischen Ereignisse einigermaßen einzufangen, werden die Annahmen zur Migration und Mortalität in einem Alters- bzw. Projektionsintervall aufgespalten. Während die Geburten genau zur Jahresmitte stattfinden, findet Migration jeweils zur Hälfte zu Beginn und am Ende eines Projektionsintervalls statt. Da alle Personen zur Jahresmitte  $x, 0$  Jahre alt sind, muss die Mortalität so aufgespalten werden, dass mit der ersten bzw. zweiten Überlebenswahrscheinlichkeit,  $p_1$  bzw.  $p_2$ , das Überleben im ersten bzw. zweiten Altersintervall,  $[x, 0; x, 5)$  bzw.  $[x, 0; x, 5)$ , also vom Jahresdurchschnittsbestand zum nächsten Jahresendbestand und vom Jahresendbestand zum nächsten Jahresdurchschnittsbestand, bestimmt werden kann (vgl. Abbildung 10.2).<sup>1</sup>

Um nun die Jahresdurchschnittsbestände und Jahresendbestände der Subpopulationen über den gesamten Projektionshorizont mit den subpopulationsspezifischen Fertilitäts- und Mortalitätsannahmen zu prognostizieren, müssen entsprechende Modellgleichungen aufgestellt werden. Da sich die Berechnungen in den einzelnen Subpopulation teilweise unterscheiden, werden sie nachfolgend, wenn nötig, für die Einheimischen, die direkten Zu- und Fortzüge wie auch für die Nachkommengenerationen der direkten Zu- und Fortzüge separat beschrieben.

In den subpopulationsspezifischen Berechnungsformeln werden Indizes verwendet, damit eine eindeutige Identifikation der zu verwendenden Parameter gewährleistet werden kann. Die Indizes  $E$ ,  $SZ$  und  $SFz$  werden für die Subpopulationen der Einheimischen, der direkten Zuzüge und der direkten Fortzüge verwendet. Für die jeweils  $n$ -te Nachkommengenerati-

---

<sup>1</sup>Im Anhang B wird die Anpassung der Reihenfolge der demographischen Ereignisse im PPPM an die Datenerfassung in Deutschland nochmal gesondert erläutert.

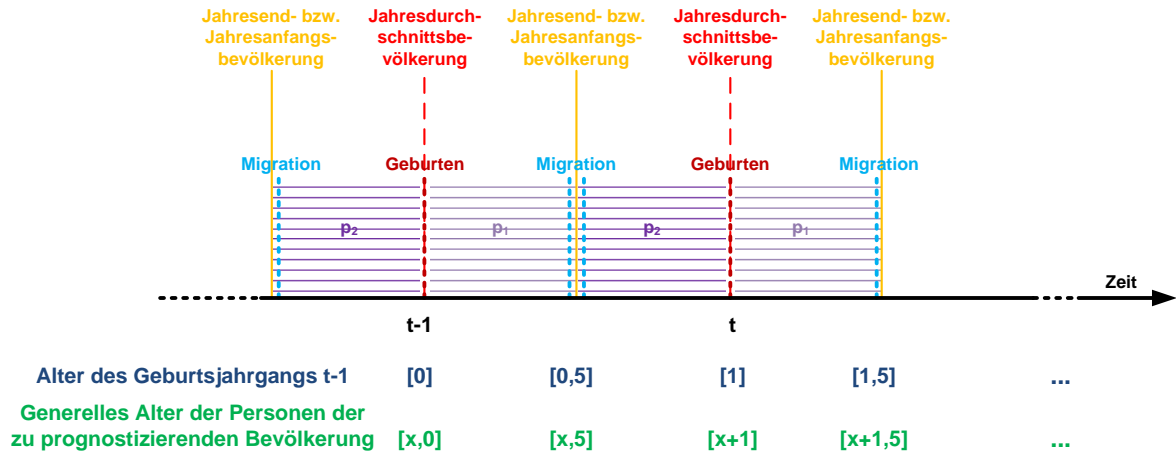


Abbildung 10.2.: Reihenfolge demographischer Ereignisse im PPPM: Geburten (dunkelrot) finden zur Jahresmitte statt, Mortalität findet mit der ersten Überlebenswahrscheinlichkeit  $p_1$  (helllila) vom Jahresdurchschnittsbestand zum Jahresendbestand und mit der zweiten Überlebenswahrscheinlichkeit  $p_2$  (lila) vom Jahresanfangsbestand zum Jahresdurchschnittsbestand statt; Migration (hellblau) findet jeweils zur Hälfte am Beginn und am Ende eines Beobachtungsjahres statt.

on der Subpopulationen der direkten Zuzüge bzw. Fortzüge werden die Indizes  $SZG(n)$  und  $SFzG(n)$  verwendet. Annahmenwerte für direkte Zu- und Fortzüge werden mit den Indizes  $Z$  und  $Fz$  gekennzeichnet. Mit den Indizes  $F$  und  $M$  können zusätzlich Frauen und Männer unterschieden werden. Grundsätzlich sind die mathematischen Berechnungsformeln für Männer und Frauen identisch bis auf wenige Ausnahmen wie die Berechnung der männlichen und weiblichen Geburten. Ansonsten unterscheiden sie sich zwar in ihren geschlechtsspezifischen Annahmenwerten, jedoch soll dies in den Indizes aus Gründen der Vereinfachung nicht weiter berücksichtigt werden. Hilfsgrößen und Ergebnisgrößen, die es in mehreren Subpopulationen gibt und deren mathematische Berechnung identisch ist, werden mit dem Index  $S$  für *Subpopulation* gekennzeichnet. Allerdings unterscheiden sich die zu prognostizierenden Werte für diese Hilfsgrößen und Modellparameter aufgrund der subpopulationsspezifischen Annahmenwerte.<sup>2</sup>

<sup>2</sup>Die hier zu beschreibenden definitorischen Zusammenhänge der Modellparameter basieren größtenteils auf dem unveröffentlichten deterministischen Bevölkerungsprognosemodell von Dinkel, wobei diese erstmals in einer einheitlichen Notation zusammengetragen und überarbeitet worden sind.

### 10.1.1. Halbjährige Überlebenswahrscheinlichkeiten

Die halbjährigen Überlebenswahrscheinlichkeiten<sup>3</sup> stellen eine Hilfsgröße zur Berechnung der Ergebnisgrößen Jahresendbestand und Jahresdurchschnittsbestand im PPPM dar. Sie geben die Wahrscheinlichkeit an, mit der eine Person das jeweils zu betrachtende halbe Jahr überlebt.

Die ersten und zweiten halbjährigen Überlebenswahrscheinlichkeiten,  $p_{1,[x,0;x,5),t}^S$  und  $p_{2,[x,5;x+1),t}^S$ , werden für alle Subpopulationen, also für die Einheimischen, die direkten Zu- und Fortzüge wie auch für jede Nachkommengeneration der direkten Zu- und Fortzüge, mit derselben Formel berechnet. Jedoch unterscheiden sich die bei ihrer Berechnung zu verwendenden, jeweils subpopulationsspezifisch anzunehmenden Werte für die Modellparameter *Anteil der Säuglingssterbefälle im ersten Halbjahr* und *Überlebende im Alter  $x$* .

Die erste, subpopulationsspezifische, halbjährige Überlebenswahrscheinlichkeit für das Altersintervall  $[0; 0,5)$  des Jahres  $t$ ,  $p_{1,[0;0,5),t}^S$ , wird zur Berechnung der jeweils subpopulationsspezifischen Jahresendbevölkerung im Alter 0,5 für ein Jahr  $t$  herangezogen. Dabei wird neben den *Überlebenden im Alter  $x$*  im Jahr  $t$ ,  $l_{x,t}^S$ , auch der Modellparameter *Anteil der Säuglingssterbefälle im ersten Halbjahr* im Jahr  $t$ ,  $\pi_{(0,5)D_0}_t^S$ , verwendet (vgl. Formel 10.1).

$$p_{1,[0;0,5),t}^S = 1 - \pi_{(0,5)D_0}_t^S \cdot \frac{l_{0,t}^S - l_{1,t}^S}{l_{0,t}^S} \quad (10.1)$$

Zur Berechnung der ersten, subpopulationsspezifischen, halbjährigen Überlebenswahrscheinlichkeit überhalb von Alter 0,5 für ein Jahr  $t$ ,  $p_{1,[x;x,5),t}^S$ , wird neben den *Überlebenden im Alter  $x$*  im Jahr  $t$  auch der Faktor 0,5 verwendet. Dieser Faktor unterstellt eine Gleichverteilung der Sterblichkeit im ganzen Altersintervall  $[x, x+1)$ , sodass gleich viele Sterbefälle in den beiden Teilintervallen,  $[x, 0; x, 5)$  und  $[x, 5; x+1)$ , auftreten (vgl. Formel 10.2).

$$p_{1,[x;x,5),t}^S = 1 - 0,5 \cdot \frac{l_{x,t}^S - l_{x+1,t}^S}{l_{x,t}^S} \quad (10.2)$$

Im Prinzip wird die erste halbjährige Überlebenswahrscheinlichkeit über die Differenz aus 1 und der ersten halbjährigen Sterbewahrscheinlichkeit ermittelt. Bei der Berechnung der ersten halbjährigen Sterbewahrscheinlichkeit für ein Alter  $x$  in einem Jahr  $t$  wird die Anzahl der

---

<sup>3</sup>Überlebenswahrscheinlichkeiten im Alter  $x$ ,  $p_x$ , geben die Wahrscheinlichkeit an, mit der eine Person das Altersintervall von  $x$  bis unter  $x+1$  überlebt. Sie berechnen sich über die Differenz aus 1 und der Sterbewahrscheinlichkeit im Alter  $x$ :  $p_x = 1 - q_x$ .

Sterbefälle aus dem ersten Halbjahr  ${}_{0,5}D_{x,t}$  durch die Gesamtzahl aller Personen zu Beginn im exakten Alter  $x$  im Jahr  $t$ ,  $l_{x,t}$ , geteilt. Die Anzahl der Sterbefälle im ersten Halbjahr wird dabei aus dem Produkt des Anteils der Sterbefälle im ersten Halbjahr und den Sterbefällen im gesamten ersten Altersjahr berechnet; für die unterste Altersstufe wird dafür der *Anteil der Säuglingssterbefälle im ersten Halbjahr* eingesetzt und für alle weiteren Altersstufen wird eine Gleichverteilung der Sterbefälle über das gesamte Altersjahr angenommen, sodass der Faktor 0,5 eingesetzt werden kann. Demzufolge können die Formeln 10.1 und 10.2 zum näheren Verständnis in die Formeln 10.3 und 10.4 umgeschrieben werden.<sup>4</sup>

$$p_{1,[0;0,5),t}^S = 1 - \frac{\pi({}_{0,5}D_0)_t^S \cdot {}_1D_{0,t}^S}{l_{0,t}^S} \quad (10.3)$$

$$p_{1,[x;x,5),t}^S = 1 - \frac{0,5 \cdot {}_1D_{x,t}^S}{l_{x,t}^S} \quad (10.4)$$

Die zweite, subpopulationsspezifische, halbjährige Überlebenswahrscheinlichkeit für das Altersintervall  $[0,5;1)$  des Jahres  $t$ ,  $p_{2,[5;1),t}$ , wird zur Berechnung der jeweils subpopulationsspezifischen Jahresdurchschnittsbevölkerung im Alter 1 für ein Jahr  $t+1$  herangezogen. Dabei wird neben den *Überlebenden im Alter  $x$  aus dem Jahr  $t$*  auch wieder der Modellparameter *Anteil der Säuglingssterbefälle im ersten Halbjahr* im Jahr  $t$  verwendet (vgl. Formel 10.5).

$$p_{2,[0,5;1),t}^S = 1 - \frac{(1 - \pi({}_{0,5}D_0)_t^S) \cdot (l_{0,t}^S - l_{1,t}^S)}{l_{0,t}^S - \pi({}_{0,5}D_0)_t^S \cdot (l_{0,t}^S - l_{1,t}^S)} \quad (10.5)$$

Zur Berechnung der zweiten, subpopulationsspezifischen, halbjährigen Überlebenswahrscheinlichkeit überhalb von Alter 1 für ein Jahr  $t$ ,  $p_{2,[x,5;x+1),t}$ , wird neben den *Überlebenden im Alter  $x$  aus dem Jahr  $t$*  auch wieder der Faktor 0,5 verwendet, der eine Gleichverteilung der Sterblichkeit im ganzen Altersintervall  $[x, x+1)$  unterstellt (vgl. Formel 10.6).

$$p_{2,[x,5;x+1),t}^S = 1 - \frac{0,5 \cdot (l_{x,t}^S - l_{x+1,t}^S)}{l_{x,t}^S - 0,5 \cdot (l_{x,t}^S - l_{x+1,t}^S)} \quad (10.6)$$

Prinzipiell wird auch die zweite halbjährige Überlebenswahrscheinlichkeit über die Differenz aus 1 und der zweiten halbjährigen Sterbewahrscheinlichkeit berechnet. Zur Berechnung der zweiten halbjährigen Sterbewahrscheinlichkeit vom Alter  $x,5$  zum Alter  $x+1$  im Jahr  $t$

<sup>4</sup>Bei den hier zu betrachtenden Sterbefällen handelt es sich um Sterbetafelsterbefälle und nicht um Sterbefallzahlen aus einer realen Bevölkerung. Aus Gründen der Vereinfachung wird das nicht zusätzlich in der Notation berücksichtigt.

wird die Anzahl der Sterbefälle aus dem zweiten Halbjahr durch die Personen zu Beginn dieses zu betrachtenden Halbjahres geteilt. Die Anzahl der Sterbefälle im zweiten Halbjahr wird über das Produkt aus dem Anteil der Sterbefälle für das zweite Halbjahr und den Sterbefällen über das gesamte Altersjahr berechnet. Der Anteil der Sterbefälle im zweiten Halbjahr wird über die Differenz aus 1 und dem Anteil der Sterbefälle im ersten Halbjahr ermittelt. Für die unterste Altersgruppe wird dafür der jeweilige Annahmenwert des Modellparameters *Anteil der Säuglingssterbefälle im ersten Halbjahr* und für alle weiteren Altersstufen wird dafür der Wert 0,5 eingesetzt. Darüber hinaus wird die Anzahl der Personen zu Beginn des zweiten Halbjahres über die Differenz aus den Personen zu Beginn des ersten Halbjahres und den Sterbefällen im ersten Halbjahr berechnet.<sup>5</sup> Somit können die Formeln 10.5 und 10.6 zum näheren Verständnis in die Formeln 10.7 und 10.8 umgeschrieben werden.

$$p_{2,[0,5;1),t}^S = 1 - \frac{(1 - \pi_{(0,5}D_0)_t^S) \cdot {}_1D_{0,t}^S}{l_{0,t}^S - \pi_{(0,5}D_0)_t^S \cdot {}_1D_{0,t}^S} \quad (10.7)$$

$$p_{2,[x,5;x+1),t}^S = 1 - \frac{0,5 \cdot {}_1D_{x,t}^S}{l_{x,t}^S - 0,5 \cdot {}_1D_{x,t}^S} \quad (10.8)$$

### 10.1.2. Jahresdurchschnittsbestände überhalb von Alter 0

Jahresdurchschnittsbestände stellen eine wesentliche Ergebnisgröße im PPPM dar auf die bei der Prognoserechnung auch immer wieder zurückgegriffen wird.

Es wird zunächst die Berechnung der Jahresdurchschnittsbestände überhalb von Alter 0 betrachtet, da die subpopulationsspezifischen Jahresdurchschnittsbestände der Frauen im reproduktiven Alter im Jahr  $t$  zur Berechnung der Geburten im selben Jahr benötigt werden (vgl. Abschnitt 10.1.3).

Die mathematischen Formeln zur Berechnung der Jahresdurchschnittsbestände unterscheiden sich zwischen den einzelnen Subpopulationen wie auch innerhalb der einzelnen Subpopulationen zwischen verschiedenen Altern.

---

<sup>5</sup>Auch hier werden wie bei den ersten, halbjährigen Überlebenswahrscheinlichkeiten Sterbetafelsterbefälle betrachtet.

### Einheimische und Nachkommengenerationen der direkten Zu- und Fortzüge

Die Berechnung der Jahresdurchschnittsbestände ist für die Einheimischen und jede der zu betrachtenden Nachkommengenerationen der direkten Zu- und Fortzüge identisch, sodass bei den Formeln stellvertretend der Index  $S$  für diese Subpopulationen verwendet wird. Die Ergebnisgrößen Jahresendbevölkerung und Jahresdurchschnittsbevölkerung, die Hilfsgröße zweite halbjährige Überlebenswahrscheinlichkeit und der optionale Modellparameter *Überlebenswahrscheinlichkeit der obersten, nach oben offenen Altersgruppe* kommen bei der Berechnung der subpopulationsspezifischen Jahresdurchschnittsbestände mit subpopulationsspezifischen Ergebnis- und Annahmewerten zum Einsatz.

Die subpopulationsspezifischen Jahresdurchschnittsbestände im Alter  $x$  im Jahr  $t$ ,  $\bar{P}_{x,t}^S$ , werden für die Alter 1 bis 94 über das Produkt aus dem Jahresendbestand der jeweiligen Subpopulation im Alter  $x - 0,5$  im Jahr  $t - 1$ ,  $P_{(x-0,5),31.12.t-1}^S$ , und der zweiten, halbjährigen Überlebenswahrscheinlichkeit im Jahr  $t - 1$ ,  $p_{2,[x-0,5;x),t-1}^S$ , berechnet (vgl. Formel 10.9).

$$\bar{P}_{x,t}^S = P_{(x-0,5),31.12.t-1}^S \cdot p_{2,[x-0,5;x),t-1}^S \quad (10.9)$$

Folglich ergibt sich die Anzahl der Personen zur Jahresmitte aus den Personen zum vorherigen Jahresende abzüglich derer, die im zweiten Altershalbjahr gestorben sind.

Die subpopulationsspezifischen Jahresdurchschnittsbestände werden für die oberste, nach oben offene Altersgruppe, wie hier beispielsweise für die 95+-Jährigen, im Jahr  $t$ ,  $\bar{P}_{95+,t}^S$ , über die Summe zweier Personengesamtheiten berechnet. Einerseits können Personen neu dazukommen, die zum Zeitpunkt des vorherigen Jahresendbestands im Alter 94,5 gewesen sind und die das halbe Jahr bis zum zu betrachtenden Jahresdurchschnittsbestand im Jahr  $t$  überleben. Um diese Personengesamtheit zu bestimmen, wird der Jahresendbestand der Personen im Alter 94,5 aus dem Jahr  $t - 1$  mit der zweiten halbjährigen Überlebenswahrscheinlichkeit vom Alter 94,5 zum Alter 95 aus dem Jahr  $t - 1$  multipliziert. Andererseits können Personen, die bereits zum Zeitpunkt des vorherigen Jahresendbestands in der Altersgruppe 95+ gewesen sind weiterhin überleben. Diese Personengesamtheit wird aus dem Produkt des Jahresendbestands der 95+-Jährigen aus dem Jahr  $t - 1$  und der Überlebenswahrscheinlichkeit der 95+-Jährigen im Jahr  $t$  berechnet. Die Überlebenswahrscheinlichkeit der 95+-Jährigen im Jahr  $t$  wird mit den subpopulationsspezifischen Annahmewerten des Modellparameters *Überlebenswahrscheinlichkeit der obersten, nach oben offenen Altersgruppe* bestimmt (vgl.

Formel 10.10 auf Seite 191).

$$\overline{P}_{95+,t}^S = P_{(94,5),31.12.t-1}^S \cdot p_{2,[94,5;95),t-1}^S + P_{(95+),31.12.t-1}^S \cdot p_{95+,t}^S \quad (10.10)$$

### Direkte Zu- und Fortzüge

Die mathematischen Formeln zur Berechnung der Jahresdurchschnittsbestände sind für die direkten Zu- und Fortzüge nahezu identisch, nur dass bei der Subpopulation der direkten Zuzüge,  $SZ$ , anzunehmende Zuzüge  $Z$  und bei der Subpopulation der direkten Fortzüge,  $SFz$ , anzunehmende Fortzüge  $Fz$  in einem Alter  $x$  in einem Jahr  $t$  betrachtet werden. Zudem unterscheiden sich die Berechnungsformeln innerhalb dieser Subpopulationen zwischen verschiedenen Altern. Aus Gründen der Übersichtlichkeit werden die Formeln für die direkten Zu- und Fortzüge separat ausgewiesen.

Die Jahresdurchschnittsbestände der Subpopulationen der direkten Zu- bzw. Fortzüge haben keine Nulljährigen, da sie mit ihren zu realisierenden Geburten jeweils die erste Nachkommengeneration aufbauen (vgl. Abschnitt 10.1.3). Außerdem müssten die Nulljährigen zur Jahresmitte der direkten Zu- und Fortzüge exakt *null* Jahre alt sein, da ansonsten die Modellannahmen zur Fertilität und damit auch zum Alter verletzt werden würden. Jedoch ist eine Wanderung im exakten Alter *null* sehr unwahrscheinlich, sodass sie im PPPM ausgeschlossen wird.

Die dennoch stattfindenden Zu- bzw. Fortzüge im Alter 0, die sich über ein ganzes Kalenderjahr  $t - 1$  erstrecken, werden im Jahresdurchschnittsbestand der direkten Zu- bzw. Fortzüge im Alter 1 im Jahr  $t$  berücksichtigt. Allerdings müssen die bis zur Mitte des Jahres  $t$  stattfindenden Säuglingssterbefälle abgezogen werden. Neben den nulljährigen kommt auch noch die Hälfte der einjährigen Zu- bzw. Fortzüge zum Jahresdurchschnittsbestand der direkten Zu- bzw. Fortzüge hinzu, die ebenfalls vom Jahresanfang bis zur Jahresmitte im Jahr  $t$  überleben müssen.

Demnach berechnen sich die Jahresdurchschnittsbestände der direkten Zu- bzw. Fortzüge im Alter 1 im Jahr  $t$ ,  $\overline{P}_{1,t}^{SZ}$  und  $\overline{P}_{1,t}^{SFz}$ , über die Summe der Zu- bzw. Fortzüge im Alter 0 im Jahr  $t - 1$  und der Hälfte der Zu- bzw. Fortzüge im Alter 1 im Jahr  $t$  multipliziert mit der zweiten subpopulationsspezifischen halbjährigen Überlebenswahrscheinlichkeit vom Alter 0, 5 zum Alter 1 aus dem Jahr  $t - 1$ . Die Formel 10.11 zeigt die Berechnung der Jahresdurch-

schnittsbestände der direkten Zuzüge im Alter 1 im Jahr  $t$  und die Formel 10.12 zeigt die Berechnung der Jahresdurchschnittsbestände der direkten Fortzüge im Alter 1 im Jahr  $t$ .

$$\bar{P}_{1,t}^{SZ} = (Z_{0,t-1} + \frac{Z_{1,t}}{2}) \cdot p_{2,[0,5;1),t-1}^{SZ} \quad (10.11)$$

$$\bar{P}_{1,t}^{SFz} = (Fz_{0,t-1} + \frac{Fz_{1,t}}{2}) \cdot p_{2,[0,5;1),t-1}^{SFz} \quad (10.12)$$

Die Jahresdurchschnittsbestände der direkten Zu- bzw. Fortzüge setzen sich für die Alter 2 bis 94 im Jahr  $t$  aus den Überlebenden des vorherigen Jahresendbestands aus dem Jahr  $t-1$  und der Hälfte der Zu- bzw. Fortzüge zum Jahresanfang im Jahr  $t$  zusammen.

Folglich berechnen sich die Jahresdurchschnittsbestände der direkten Zu- bzw. Fortzüge im Alter  $x$  im Jahr  $t$ ,  $\bar{P}_{x,t}^{SZ}$  und  $\bar{P}_{x,t}^{SFz}$ , für die Alter 2 bis 94 aus der Summe der Hälfte der Zu- bzw. Fortzüge im Alter  $x$  im Jahr  $t$  und dem Jahresendbestand der direkten Zu- bzw. Fortzüge im Alter  $x-0,5$  im Jahr  $t-1$ ,  $P_{(x-0,5),31.12.t-1}^{SZ}$  und  $P_{(x-0,5),31.12.t-1}^{SFz}$ , multipliziert mit der zweiten subpopulationsspezifischen halbjährigen Überlebenswahrscheinlichkeit vom Alter  $x-0,5$  zum Alter  $x$  aus dem Jahr  $t-1$  (vgl. Formeln 10.13 und 10.13).

$$\bar{P}_{x,t}^{SZ} = (\frac{Z_{x,t}}{2} + P_{(x-0,5),31.12.t-1}^{SZ}) \cdot p_{2,[x-0,5;x),t-1}^{SZ} \quad (10.13)$$

$$\bar{P}_{x,t}^{SFz} = (\frac{Fz_{x,t}}{2} + P_{(x-0,5),31.12.t-1}^{SFz}) \cdot p_{2,[x-0,5;x),t-1}^{SFz} \quad (10.14)$$

Die Jahresdurchschnittsbestände der direkten Zu- bzw. Fortzüge für die 95+-Jährigen im Jahr  $t$ ,  $\bar{P}_{95+,t}^{SZ}$  und  $\bar{P}_{95+,t}^{SFz}$ , setzen sich jeweils aus drei Personengesamtheiten zusammen. Neben den Überlebenden des vorherigen Jahresendbestands im Alter 94,5 und 95+ können nun auch Überlebende neuer Zu- bzw. Fortzüge im Alter 95+ vom Jahresanfang dazukommen. Die Überlebenden des vorherigen Jahresendbestands berechnen sich über das Produkt aus den Beständen in den Altern 94,5 und 95+ aus dem Jahr  $t-1$  und den entsprechenden Überlebenswahrscheinlichkeiten,  $p_{2,[94,5;95),t-1}^S$  und  $p_{95+,t}^S$ . Wohingegen die Überlebenden der Hälfte der Zu- bzw. Fortzüge über das Produkt aus der Hälfte der Zu- bzw. Fortzüge im Alter 95+ im Jahr  $t$  und der subpopulationsspezifischen Überlebenswahrscheinlichkeit der 95+-Jährigen bestimmt werden (vgl. Formeln 10.15 und 10.16).

$$\bar{P}_{95+,t}^{SZ} = P_{(94,5),31.12.t-1}^{SZ} \cdot p_{2,[94,5;95),t-1}^S + (\frac{Z_{95+,t}}{2} + P_{(95+),31.12.t-1}^{SZ} \cdot p_{95+,t}^S) \quad (10.15)$$



$$\overline{P}_{95+,t}^{SFz} = P_{(94,5),31.12.t-1}^{SFz} \cdot p_{2,[94,5;95),t-1}^{SFz} + \left( \frac{Fz_{95+,t}}{2} + P_{(95+),31.12.t-1}^{SFz} \cdot p_{95+,t}^{SFz} \right) \quad (10.16)$$

### 10.1.3. Geburten

Nachdem die Jahresdurchschnittsbestände der Frauen im reproduktiven Alter für die Subpopulationen der Einheimischen, der direkten Zu- und Fortzüge wie auch für jede Nachkommengeneration der direkten Zu- und Fortzüge für das Jahr  $t$  mit den Formeln 10.9, 10.13 und 10.14 prognostiziert worden sind, können auch die subpopulationsspezifischen Geburten für das Jahr  $t$  berechnet werden. Dazu werden neben den Jahresdurchschnittsbeständen der Frauen auch die subpopulationsspezifischen Annahmenwerte der Modellparameter *altersspezifische Fertilitätsraten* und *Geschlechteranteil bei Geburt* benötigt. Jedoch unterscheiden sich die Berechnungen der Nulljährigen für die einzelnen Subpopulationen.

Im PPPM hängt die Zahl der Geburten ausschließlich vom Jahresdurchschnittsbestand der Frauen und ihren altersspezifischen Fertilitätsraten ab. Dies liegt darin begründet, dass potentiellen Müttern aufgrund der Schwangerschaftsdauer nicht mehr als eine Geburt eines oder mehrerer neugeborener Kinder in einem Jahr mit relativ großer Sicherheit zugeordnet werden kann. Im Gegensatz dazu können Männer nicht nur mehrere Kinder in einem Jahr, sondern zudem auch über eine längere Altersspanne (er)zeugen. Da die Zuordnung von Kindern demnach bei Männern mit einer größeren Unsicherheit verbunden ist, wird die Fertilität im PPPM allein den Frauen zugeschrieben.

#### Einheimische

Zur Berechnung der Geburten der Einheimischen  $B_t^E$  wird zunächst ein Summenprodukt aus der durchschnittlichen Zahl der einheimischen Frauen im reproduktiven Alter  $x^6$  und den entsprechenden *altersspezifischen Fertilitätsraten* berechnet. Anschließend wird dieses Summenprodukt mit dem jeweiligen *Geschlechteranteil bei Geburt* multipliziert, um auch die Anzahl männlicher und weiblicher einheimischer Geburten,  $B_t^{E,M}$  und  $B_t^{E,F}$ , im Jahr  $t$  zu bestimmen. Die weiblichen einheimischen Geburten des Jahres  $t$  werden im PPPM mit der

---

<sup>6</sup>Als reproduktive Altersstufen gelten für gewöhnlich die Alter 15 bis 50. Im PPPM sollen jedoch auch Schwangerschaften jüngerer (und älterer) Frauen berücksichtigt werden können, sodass Annahmen zur Fertilität in jeder Altersstufe getroffen werden müssen. Selbstverständlich können z. B. einjährige Frauen noch kein Kind bekommen, sodass ihre entsprechende Fertilitätsrate *null* ist. Jedoch wird somit im PPPM weder eine untere noch eine obere Grenze (für die Fertilität) fest vorgegeben.

Formel 10.17 berechnet, wohingegen die männlichen einheimischen Geburten des Jahres  $t$  mit der Formel 10.18 berechnet werden.

$$B_t^{E,F} = \left( \sum_{x=1}^{50} \bar{P}_{x,t}^{E,F} \cdot f_{x,t}^E \right) \cdot \pi(P_0^{E,F}) \quad (10.17)$$

$$B_t^{E,M} = \left( \sum_{x=1}^{50} \bar{P}_{x,t}^{E,F} \cdot f_{x,t}^E \right) \cdot \pi(P_0^{E,M}) \quad (10.18)$$

### Direkte Zu- und Fortzüge

Die Geburten der Jahresdurchschnittsbestände der direkten Zu- und Fortzüge werden jeweils in der ersten Nachkommengeneration der Zu- und Fortzüge ausgewiesen, sodass der Bestand der Nulljährigen bei den Jahresdurchschnittsbeständen der direkten Zu- und Fortzüge immer *null* ist. Folglich bauen sich diese beiden Subpopulationen allein aufgrund der direkt stattfindenden Zu- und Fortzüge auf (vgl. Abschnitt 10.1.2).

### Nachkommengenerationen der direkten Zu- und Fortzüge

Die Berechnung der Geburten für die Nachkommengenerationen der direkten Zu- bzw. Fortzüge weicht nur in einem wesentlichen Punkt von der Berechnung der Geburten für die Einheimischen ab. Anstatt der Frauen im reproduktiven Alter der jeweiligen Nachkommengeneration werden die Frauen im reproduktiven Alter aus der jeweils vorherigen Nachkommengeneration als Mütter eingesetzt. Folglich bauen sich die Nachkommengenerationen sukzessive auf, wobei jede Nachkommengeneration von der vorherigen Nachkommengeneration gebildet wird. Die erste Nachkommengeneration stellt zudem jeweils einen Sonderfall dar, da hier die Frauen im reproduktiven Alter der Subpopulationen der direkten Zu- bzw. Fortzüge als Mütter eingesetzt werden.

Demzufolge wird bei der Berechnung der männlichen und weiblichen Geburten für die erste Nachkommengeneration der direkten Zu- bzw. Fortzüge im Jahr  $t$ ,  $B_t^{SZG(1),M}$ ,  $B_t^{SZG(1),F}$ ,  $B_t^{SFzG(1),M}$  und  $B_t^{SFzG(1),F}$ , das Summenprodukt aus den Jahresdurchschnittsbeständen der Frauen der direkten Zu- bzw. Fortzüge im Alter  $x$  im Jahr  $t$  und den entsprechenden *altersspezifischen Fertilitätsraten* mit dem jeweiligen *Geschlechteranteil bei Geburt* multipliziert (vgl. Formeln 10.19, 10.20, 10.21 und 10.22).

$$B_t^{SZG(1),F} = \left( \sum_{x=1}^{50} \bar{P}_{x,t}^{SZ,F} \cdot f_{x,t}^{SZ} \right) \cdot \pi(P_0^{SZ,F}) \quad (10.19)$$

$$B_t^{SZG(1),M} = \left( \sum_{x=1}^{50} \bar{P}_{x,t}^{SZ,F} \cdot f_{x,t}^{SZ} \right) \cdot \pi(P_0^{SZ,M}) \quad (10.20)$$

$$B_t^{SFzG(1),F} = \left( \sum_{x=1}^{50} \bar{P}_{x,t}^{SFz,F} \cdot f_{x,t}^{SFz} \right) \cdot \pi(P_0^{SFz,F}) \quad (10.21)$$

$$B_t^{SFzG(1),M} = \left( \sum_{x=1}^{50} \bar{P}_{x,t}^{SFz,F} \cdot f_{x,t}^{SFz} \right) \cdot \pi(P_0^{SFz,M}) \quad (10.22)$$

Zur Berechnung der männlichen und weiblichen Geburten ab der zweiten Nachkommengeneration der direkten Zu- bzw. Fortzüge im Jahr  $t$ ,  $B_t^{SZGn,M}$  und  $B_t^{SZGn,F}$ , wird das Summenprodukt aus den Jahresdurchschnittsbeständen der Frauen der vorherigen Nachkommengeneration im Alter  $x$  im Jahr  $t$  und den entsprechenden *altersspezifischen Fertilitätsraten* mit dem jeweiligen *Geschlechteranteil bei Geburt* multipliziert (vgl. Formeln 10.23, 10.24, 10.25 und 10.26).

$$B_t^{SZG(n>1),F} = \left( \sum_{x=1}^{50} \bar{P}_{x,t}^{SZG(n-1),F} \cdot f_{x,t}^{SZG(n-1)} \right) \cdot \pi(P_0^{SZG(n-1),F}) \quad (10.23)$$

$$B_t^{SZG(n>1),M} = \left( \sum_{x=1}^{50} \bar{P}_{x,t}^{SZG(n-1),F} \cdot f_{x,t}^{SZG(n-1)} \right) \cdot \pi(P_0^{SZG(n-1),M}) \quad (10.24)$$

$$B_t^{SFzG(n>1),F} = \left( \sum_{x=1}^{50} \bar{P}_{x,t}^{SFzG(n-1),F} \cdot f_{x,t}^{SFzG(n-1)} \right) \cdot \pi(P_0^{SFzG(n-1),F}) \quad (10.25)$$

$$B_t^{SFzG(n>1),M} = \left( \sum_{x=1}^{50} \bar{P}_{x,t}^{SFzG(n-1),F} \cdot f_{x,t}^{SFzG(n-1)} \right) \cdot \pi(P_0^{SFzG(n-1),M}) \quad (10.26)$$

#### 10.1.4. Jahresendbestände

Die Jahresendbestände stellen wie die Jahresdurchschnittsbestände eine wesentliche Ergebnisgröße im PPPM dar auf die bei der Berechnung der Prognose zurückgegriffen wird.

Die mathematischen Formeln zur Berechnung der Jahresendbestände unterscheiden sich zwischen den einzelnen Subpopulationen wie auch innerhalb der Subpopulationen zwischen verschiedenen Altern.

### Einheimische und Nachkommengenerationen der direkten Zu- und Fortzüge

Die Berechnung der Jahresendbestände im Alter  $x,5$  im Jahr  $t$ ,  $P_{x,5;31.12.t}$ , ist für die Subpopulationen der Einheimischen und der Nachkommengenerationen der direkten Zu- und Fortzüge identisch, sodass ersatzweise der Index  $S$  für diese Subpopulationen in den Formeln verwendet wird.

Die Jahresendbestände sind quasi die Überlebenden des vorherigen Jahresdurchschnittsbestands. Folglich werden die subpopulationsspezifischen Jahresendbestände der Personen in den Altern  $0,5$  bis  $94,5$  im Jahr  $t$  über das Produkt aus den jeweiligen subpopulationsspezifischen Jahresdurchschnittsbeständen im Alter  $x$  im Jahr  $t$  und den jeweiligen ersten, subpopulationsspezifischen, halbjährigen Überlebenswahrscheinlichkeiten vom Alter  $x,0$  zum Alter  $x,5$  im Jahr  $t$  berechnet (vgl. Formel 10.27).

$$P_{x,5;31.12.t}^S = \bar{P}_{x,t}^S \cdot p_{1,[x;5),t}^S \quad (10.27)$$

Die subpopulationsspezifischen Jahresendbestände der 95+-Jährigen entsprechen den Überlebenden des vorherigen Jahresendbestands der 95+-Jährigen. Demzufolge werden die subpopulationsspezifischen Jahresendbestände im Alter 95+ im Jahr  $t$ ,  $P_{95+;31.12.t}^S$ , über das Produkt aus den subpopulationsspezifischen Jahrdurchschnittsbeständen im Alter 95+ im Jahr  $t$  und den entsprechenden Annahmenwerten des Modellparameters *Überlebenswahrscheinlichkeit der obersten, nach oben offenen Altersgruppe* im Jahr  $t$  berechnet (vgl. Formel 10.28).

$$P_{95+;31.12.t}^S = \bar{P}_{95+,t}^S \cdot p_{95+,t}^S \quad (10.28)$$

### Direkte Zu- und Fortzüge

Die Berechnung der Jahresendbestände der Subpopulationen der direkten Zu- und Fortzüge unterscheidet sich durch die Verwendung anzunehmender Zuzüge  $Z$  und anzunehmender Fortzüge  $Fz$ . Außerdem unterscheiden sich die Berechnungsformeln innerhalb der jeweiligen Subpopulation zwischen verschiedenen Altern. Aus Gründen der Übersichtlichkeit werden diese mathematischen Berechnungsformeln wie bei den Jahresdurchschnittsbeständen für beide Subpopulationen separat ausgewiesen.

Generell werden die Jahresendbestände der direkten Zu- bzw. Fortzüge im Alter  $0,5$  im

Jahr  $t$ ,  $P_{0,5;31.12.t}^{SZ}$  und  $P_{0,5;31.12.t}^{SFz}$ , aus den Überlebenden des entsprechenden, vorherigen Jahresdurchschnittsbestands im Alter 0 im Jahr  $t$  und den anzunehmenden Zu- bzw. Fortzügen im Alter 0 bestimmt. Da es im PPPM keine Nulljährigen im Jahresdurchschnittsbestand der direkten Zu- und Fortzüge gibt (vgl. Abschnitt 10.1.3), entsprechen die Jahresendbestände im Alter 0,5 im Jahr  $t$  im Prinzip den anzunehmenden Zu- bzw. Fortzügen im Alter 0 im Jahr  $t$  (vgl. Formeln 10.29 und 10.30).

$$P_{0,5;31.12.t}^{SZ} = \bar{P}_{0,t}^{SZ} \cdot p_{1,[0;0,5),t}^{SZ} + Z_{0,t} \quad (10.29)$$

$$P_{0,5;31.12.t}^{SFz} = \bar{P}_{0,t}^{SFz} \cdot p_{1,[0;0,5),t}^{SFz} + F_{0,t} \quad (10.30)$$

Die Jahresendbestände der direkten Zu- bzw. Fortzüge im Alter  $x, 5$  im Jahr  $t$ ,  $P_{x,5;31.12.t}^{SZ}$  und  $P_{x,5;31.12.t}^{SFz}$ , entsprechen für die Alter 1,5 bis 94,5 im Wesentlichen den Überlebenden des entsprechenden Jahresdurchschnittsbestands im Alter  $x$  im Jahr  $t$  und der Hälfte der Zu- bzw. Fortzüge im Alter  $x$  im Jahr  $t$ . Zur Berechnung der Überlebenden des entsprechenden Jahresdurchschnittsbestands im Alter  $x$  im Jahr  $t$  wird die entsprechende erste, halbjährige Überlebenswahrscheinlichkeit vom Alter  $x$  zum Alter  $x, 5$  aus dem Jahr  $t$  verwendet (vgl. Formeln 10.31 und 10.32).

$$P_{x,5;31.12.t}^{SZ} = \bar{P}_{x,t}^{SZ} \cdot p_{1,[x;x,5),t}^{SZ} + \frac{Z_{x,t}}{2} \quad (10.31)$$

$$P_{x,5;31.12.t}^{SFz} = \bar{P}_{x,t}^{SFz} \cdot p_{1,[x;x,5),t}^{SFz} + \frac{F_{x,t}}{2} \quad (10.32)$$

Die subpopulationsspezifischen Jahresendbestände der direkten Zu- bzw. Fortzüge werden im Alter 95+ im Jahr  $t$ ,  $P_{95+;31.12.t}^{SZ}$  und  $P_{95+;31.12.t}^{SFz}$ , über das Produkt des entsprechenden Jahresdurchschnittsbestands im Alter 95+ im Jahr  $t$  und der entsprechenden anzunehmenden Überlebenswahrscheinlichkeit der 95+-Jährigen im Jahr  $t$  berechnet (vgl. Formeln 10.33 und 10.34).

$$P_{95+;31.12.t}^{SZ} = \bar{P}_{95+,t}^{SZ} \cdot p_{95+,t}^{SZ} \quad (10.33)$$

$$P_{95+;31.12.t}^{SFz} = \bar{P}_{95+,t}^{SFz} \cdot p_{95+,t}^{SFz} \quad (10.34)$$

Demzufolge entsprechen die Jahresendbestände der Subpopulationen der direkten Zu- bzw. Fortzüge im Alter 95+ den Überlebenden des vorherigen Jahresdurchschnittsbestands im Alter 95+.

### 10.1.5. Zusammenfassung

Mit diesen deterministischen Modellgleichungen können die wesentlichen Ergebnisgrößen, also die Jahresdurchschnitts- und die Jahresendbestände jeder Subpopulation nach Einzelalter, berechnet werden. (Die Tabellen 10.1 bis 10.4 ordnen die einzelnen Formeln den entsprechenden Subpopulationen und Altern nochmal überblicksartig zu.) Allerdings kann bei diesem grundlegenden deterministischen Prognosemodell immer nur ein Annahmenverlauf pro Modellparameter berücksichtigt werden, sodass auch nur ein Ergebnisverlauf pro Ergebnisgröße herauskommt. Um nun jedoch das potentielle Entwicklungsspektrum der Modellparameter wie auch der Ergebnisgrößen (möglichst vollständig) einzufangen, müssen mehrere Annahmenverläufe pro Modellparameter in einer probabilistischen Erweiterung des Grundmodells berücksichtigt werden, sodass auch eine Ergebnisverteilung für jede Ergebnisgröße resultiert. Welche Aspekte bei der Berechnung einer Ergebnisverteilung (allgemein und im PPPM) zu beachten sind, wird daher im nächsten Abschnitt beschrieben.

	<b>0</b>	<b>1 - 94</b>	<b>95+</b>
<b>E,F</b>	10.17	10.9	10.10
<b>E,M</b>	10.18	10.9	10.10
<b>SZG(1),F</b>	10.19	10.9	10.10
<b>SZG(1),M</b>	10.20	10.9	10.10
<b>SFzG(1),F</b>	10.21	10.9	10.10
<b>SFzG(1),M</b>	10.22	10.9	10.10
<b>SZG(n),F</b>	10.23	10.9	10.10
<b>SZG(n),M</b>	10.24	10.9	10.10
<b>SFzG(n),F</b>	10.25	10.9	10.10
<b>SFzG(n),M</b>	10.26	10.9	10.10

Tabelle 10.1.: Überblick über die Formeln zur Berechnung der Jahresdurchschnittsbestände für die Einheimischen und die Nachkommengenerationen der direkten Zu- und Fortzüge

	1	2 - 94	95+
<b>SZ,F</b>	10.11	10.13	10.15
<b>SZ,M</b>	10.11	10.13	10.15
<b>SFz,F</b>	10.12	10.14	10.16
<b>SFz,M</b>	10.12	10.14	10.16

Tabelle 10.2.: Überblick über die Formeln zur Berechnung der Jahresdurchschnittsbestände für die direkten Zu- und Fortzüge

	0,5 - 94,5	95+
<b>E,F</b>	10.27	10.28
<b>E,M</b>	10.27	10.28
<b>SZG(n),F</b>	10.27	10.28
<b>SZG(n),M</b>	10.27	10.28
<b>SFzG(n),F</b>	10.27	10.28
<b>SFzG(n),M</b>	10.27	10.28

Tabelle 10.3.: Überblick über die Formeln zur Berechnung der Jahresendbestände für die Einheimischen und die Nachkommengenerationen der direkten Zu- und Fortzüge

	0,5	1,5 - 94,5	95+
<b>SZ,F</b>	10.29	10.31	10.33
<b>SZ,M</b>	10.29	10.31	10.33
<b>SFz,F</b>	10.30	10.32	10.34
<b>SFz,M</b>	10.30	10.32	10.34

Tabelle 10.4.: Überblick über die Formeln zur Berechnung der Jahresendbestände für die direkten Zu- und Fortzüge

## 10.2. Berechnung *mehrerer* Annahmenverlaufskombinationen bzw. einer Ergebnisverteilung

Das PPPM ist ein probabilistisches Makromodell, dass die Einheimischen, die direkten Zu- und Fortzüge wie auch deren Nachkommengenerationen mit separaten Annahmen zur Fertilität und Mortalität prognostiziert. Dafür gibt es zur Berechnung wesentlicher Ergebnisgrößen Modellgleichungen (mit einer Reihenfolge der demographischen Ereignisse) fest vor.

Darüber hinaus können im PPPM weitere Einstellungen zur Generierung einer Annahmenverteilung für jeden Modellparameter, zur Kombination der Annahmenverläufe in einem Prognosedurchlauf wie auch zur simulativen oder (quasi-)vollständigen Berechnung vorgenommen werden, um jeweils die Erwartungen eines Prognostikers bzw. die spezifischen Anforderungen einer Bevölkerungsprognose flexibel zu erfüllen.

### 10.2.1. Erstellung einer Annahmenverteilung für jeden Modellparameter

Die zukünftige Entwicklung eines Modellparameters ist mehr oder weniger mit Unsicherheit behaftet. Folglich gibt es meist nicht nur eine, sondern mehrere potentielle Entwicklungen pro Modellparameter, die auch nicht nur mit einem, sondern nur mit mehreren Annahmenverläufen eingefangen werden können.

Bei der Erstellung einer Annahmenverteilung für einen Modellparameter kann ein Prognostiker im PPPM unterschiedliche Informationsquellen wie die vergangene Entwicklung, erklärende Theorien und Expertenmeinung heranziehen. Deuten diese Informationsquellen auf ähnliche zukünftige Entwicklungen für einen Modellparameter hin, ist die zu generierende Annahmenverteilung relativ homogen, sodass sich die potentiellen Annahmenverläufe kaum im Niveau und im Verlauf unterscheiden. Für diesen Fall bietet das PPPM die Möglichkeit, entweder einzelne, leicht voneinander abweichende Annahmenverläufe oder aber (ähnlich wie bei anderen probabilistischen Bevölkerungsprognosemodellen) einen *repräsentativen* Annahmenverlauf pro Modellparameter einzugeben. Ein repräsentativer Annahmenverlauf gibt im PPPM das Niveau und den Verlauf der Annahmenwerte (eines Modellparameters) über den Projektionshorizont richtungsweisend vor. Wird dann ein repräsentativer Annahmenverlauf (in einer simulativen Ausführung) zufällig gezogen, geht eine zufällige (proportionale) Variation der repräsentativen Annahmenwerte, bei der die repräsentative Altersstruktur in jedem Projektionsintervall erhalten bleibt, in die Berechnung der zukünftigen Bevölkerung ein. Eine zufällige (proportionale) Variation kann somit das Niveau und den Verlauf (in einem vom Prognostiker zu bestimmenden Ausmaß), nicht aber die Altersstruktur eines repräsentativen Annahmenverlaufs verändern, sodass die Annahmenwerte immer theoriekonsistent bzw. demographisch begründbar sind. Demzufolge kann trotz der Eingabe nur eines Annahmenverlaufs pro Modellparameter eine probabilistische Bevölkerungsprognose erstellt werden, die dennoch zufällige Abweichungen und damit die Prognoseunsicherheit berücksichtigt.



Deuten die unterschiedlichen Informationsquellen allerdings auf sich (stark) im Niveau und im Verlauf voneinander abweichende Annahmenverläufe für einen Modellparameter hin, so ist, wenn der Prognostiker keine dieser Entwicklungen mit Sicherheit ausschließen kann, die zu generierende Annahmenverteilung relativ heterogen. In diesem Fall bietet das PPPM (im Gegensatz zu anderen probabilistischen Bevölkerungsprognosemodellen) die Möglichkeit, einzelne, z. T. stark voneinander abweichende Annahmenverläufe pro Modellparameter zu berücksichtigen. Jeder dieser Annahmenverläufe kann zudem noch variabel als repräsentativer Annahmenverlauf eingestellt werden, sodass zusätzlich zufällige (proportionale) Variationen der richtungsweisenden Annahmenwerte in die Prognose eingehen.

Darüber hinaus wird die flexible Erstellung einer heterogenen Annahmenverteilung für einen Modellparameter im PPPM damit unterstützt, dass weder die Generierung der Annahmenverläufe noch die Zuordnung der Eintrittswahrscheinlichkeiten durch die Vorgabe einer Methode bzw. einer Verteilungsform eingeschränkt ist. Im PPPM können daher für jeden Modellparameter unbegrenzt viele Annahmenverläufe mit verschiedenen Methoden extern erzeugt werden, die (, wenn nötig, sogar deutlich) im Niveau, im Verlauf wie auch in der erwarteten, subjektiv bestimmbaren Eintrittswahrscheinlichkeit voneinander abweichen.<sup>7</sup>

Folglich kann das PPPM die Vielfalt der potentiellen Entwicklungsspektren aller Modellparameter flexibel einfangen, indem für jeden Modellparameter unbegrenzt viele Annahmenverläufe eingegeben werden können, die im Niveau und Verlauf, in der zufälligen Variabilität wie auch in der erwarteten Eintrittswahrscheinlichkeit frei vom Prognostiker bestimmt werden können. Dadurch erreicht das PPPM (im Gegensatz zu anderen probabilistischen Bevölkerungsprognoseansätzen) eine enorme Flexibilität bei der Generierung und Gewichtung der Annahmenverläufe für jeden Modellparameter, sodass damit die *wahre* Wahrscheinlichkeitsverteilung (vgl. Abschnitt 8.1.2 ab Seite 155) ohne einschränkende Vorgaben variabler approximiert werden kann. Folglich können mit dem PPPM vielfältigere Annahmenverteilungen für einen Modellparameter erzeugt werden, sodass damit, gerade bei besonders unsicheren

---

<sup>7</sup>Ohne diese Flexibilität wäre es z. B. unmöglich, zwei heterogenen Annahmenverläufen am unteren und oberen Ende des potentiellen Entwicklungsspektrums dieselbe höchste Eintrittswahrscheinlichkeit zuzuordnen, wenn eine modell-basierte Methode vorsähe, verschiedene Annahmenverläufe aus zufällig normalverteilten Abweichungen um einen vorgegebenen mittleren Annahmenverlauf zu generieren. Solche Vorgaben führen zwangsläufig zu einer Einschränkung der Annahmenverteilung hinsichtlich des Niveaus und Verlaufs wie auch der Eintrittswahrscheinlichkeit.

Prognosen, die Prognoseunsicherheit umfassender eingefangen und damit letztendlich die Prognosegenauigkeit gesteigert werden kann.

### 10.2.2. Kombination der Annahmenverläufe im Open Type und Limited Type

Pro Prognosedurchlauf wird aus der Kombination eines Annahmenverlaufs für jeden Modellparameter ein Ergebnisverlauf für jede Ergebnisgröße berechnet. Wenn nun aber mehrere Annahmenverläufe pro Modellparameter unterschiedliche Informationsquellen bzw. Theorien widerspiegeln, hängt es von den Erwartungen eines Prognostikers ab, ob alle oder nur ein Teil aller überhaupt möglichen Annahmenverlaufskombinationen plausibel sind. Sobald unplausible Kombinationen im *Open Type* entstehen, sollten diese gezielt von einem Prognostiker im *Limited Type* (durch die Einführung einer entsprechenden Ordnungsstruktur) ausgeschlossen werden können; da die Definition unplausibler Kombinationen nicht fix, sondern in Abhängigkeit von den jeweiligen Erwartungen eines Prognostikers (von Prognose zu Prognose) variieren kann, sollte daher auch die Ordnungsstruktur entsprechend flexibel sein.

#### Open Type

Im *Open Type* können die Annahmenverläufe aller Modellparameter uneingeschränkt miteinander kombiniert werden, sodass einerseits die Menge aller überhaupt möglichen Annahmenverlaufskombinationen sehr groß ist und andererseits die Plausibilität der Prognoseergebnisse beeinträchtigt werden kann.

Die Anzahl aller überhaupt möglichen Annahmenverlaufskombinationen steigt im *Open Type* übermäßig stark mit der Anzahl der Annahmenverläufe pro Modellparameter. So gibt es bereits  $4^{123}$  bzw.  $2^{2^{123}}$  Annahmenverlaufskombinationen, wenn lediglich 4 Annahmenverläufe für jeden der 123 Modellparameter eingegeben werden und jeweils 5 Nachkommengenerationen für die Immigranten und Emigranten berücksichtigt werden (vgl. Abschnitt 10.3.1). Um trotz dieser *kombinatorischen Explosion* die Prognoseunsicherheit bzw. alle potentiellen Entwicklungen einzufangen, ist eine Ergebnisverteilung fast nur simulativ zu bestimmen (vgl. Abschnitt 10.3.2).

Zudem kann die Plausibilität der Prognoseergebnisse im *Open Type* beeinträchtigt werden, wenn in einem Prognosedurchlauf durch die uneingeschränkte Kombinationsmöglichkeit (Theorie- und/oder Erwartungs-)inkonsistente Annahmenverläufe verschiedener Modellpara-

meter zufällig gezogen und miteinander kombiniert werden. Eine *unplausible Kombination* bestünde beispielsweise in der Kombination von Annahmenverläufen mit sehr hoher und sehr niedriger Fertilität, obwohl ein Prognostiker ausschließlich ein ähnlich hohes Fertilitätsniveau für alle Subpopulationen als wahrscheinlich bzw. plausibel einstuft.

### Limited Type

Die Nachteile des *Open Type* sollen im *Limited Type* durch die Einführung einer zweistufigen Ordnungsstruktur, bestehend aus *Settypen* und *Sets*, behoben werden.

In *Settypen* kann ein Prognostiker Modellparameter variabel zusammenschließen, für die er oder sie eine relationale bzw. abhängige Entwicklung in der Zukunft erwartet. Generell bestehen im PPPM keinerlei Beschränkungen bei dem Zusammenschluss der Modellparameter in Settypen, allerdings kann jeder Modellparameter immer nur in einem Settyp vorkommen. So können beispielsweise Modellparameter der Fertilität und Mortalität in einem oder auch in separaten Settypen zusammengeschlossen werden. Für jeden erstellten Settyp können dann verschiedene *Sets* definiert werden, wobei jedes Set ausschließlich konsistente Annahmenverläufe für alle Modellparameter dieses Settyps vereint. Dabei muss jedes Set mindestens einen Annahmenverlauf für jeden Modellparameter beinhalten. Bei einem Prognosedurchlauf mit dem *Limited Type* wird dann zunächst aus jedem Settyp ein Set zufällig gezogen, aus dem dann im Anschluss für jeden (im Settyp eingebundenen) Modellparameter ein Annahmenverlauf zufällig gezogen wird.

Die Einführung der flexiblen zweistufigen Ordnungsstruktur mit Settypen und Sets ermöglicht sowohl die (Theorie- oder/und Erwartungs-)konsistente Verbindung von Annahmenverläufen *eines* Modellparameters über mehrere Subpopulationen als auch die (Theorie- oder/und Erwartungs-)konsistente Verbindung von Annahmenverläufen *verschiedener* Modellparameter (über mehrere Subpopulationen). So können sowohl separate Anpassungsprozesse für die Fertilität und Mortalität von den direkten Zuzügen an das jeweilige Niveau der Einheimischen über die einzelnen Nachkommengenerationen als auch abhängige Entwicklungen zwischen der Mortalität und Fertilität (z. B. in Anlehnung an die Theorie des Parallelismus von Bertillon (vgl. Abschnitt 4.1 ab Seite 52)) modelliert werden.

Demzufolge ermöglicht die Einführung der Settypen und Sets im *Limited Type* den gezielten Ausschluss unplausibler Kombinationen, sodass nur noch plausible Annahmenverläufe

über alle Modellparameter in einem Prognosedurchlauf kombiniert werden. Dadurch kann die Menge aller möglichen Annahmenverlaufskombinationen stark reduziert und die Berechnung der Ergebnisverteilung vereinfacht werden. Dieser Aspekt wird nachfolgend durch die kombinatorische Betrachtung in Abschnitt 10.3.1 deutlich.

### 10.3. Berechnung der Ergebnisverteilung

Bevor in Abschnitt 10.3.2 konkrete Verfahren zur Berechnung bzw. Annäherung der Ergebnisverteilung einer PPPM-Prognose vorgestellt werden, sollen zunächst Konzepte wie *Settyp* und *Set*, die den *Open Type* des PPPM vom *Limited Type* unterscheiden (vgl. Abschnitt 10.2.2), formal definiert werden. Die kombinatorische Betrachtung ermöglicht dabei den Vergleich der Annahmenverlaufskombinationen des Open Type und Limited Type des PPPM.

#### 10.3.1. Kombinatorische Betrachtung — Eintrittswahrscheinlichkeiten und Anzahl der Annahmenverlaufskombinationen

Sei  $\mathbb{P} = \{p_1, \dots, p_n\}$  die Menge aller Modellparameter für eine Prognose mit dem PPPM. Diese wird in paarweise disjunkte Mengen  $P_1, \dots, P_k$  zerlegt:

$$\forall P_i, P_j \subseteq \mathbb{P} : i = j \vee P_i \cap P_j = \emptyset$$

Jede Parametermenge  $P_i$ ,  $i = 1, \dots, k$ , stellt einen *Settyp* dar, dessen *Sets* in der Menge  $S_i = \{\sigma_1, \dots, \sigma_u\}$  definiert sind. Jedes *Set*  $\sigma$  eines *Settyps*  $S_i$  sei ein Tupel  $(pr_\sigma, A_\sigma)$ , wobei  $pr_\sigma \in (0, 1]$  die Eintrittswahrscheinlichkeit des Sets und  $A_\sigma = \{A_\sigma^p | p \in P_i\}$  die Menge aller für dieses *Set* definierten Annahmenverlaufsmengen ist. Dabei wird für jeden Parameter  $p \in P_i$  eine nichtleere Annahmenverlaufsmenge  $A_\sigma^p$  definiert, dass heißt es gilt  $|A_\sigma| = |P_i|$  und  $\forall A \in A_\sigma : |A| > 0$ .

Sei  $\mathbb{S} = \{S_1, \dots, S_k\}$  die Menge aller Setmengen  $S_i$  der jeweiligen *Settypen*  $P_i$ . Für jeden *Settyp* muss mindestens ein *Set* in der jeweiligen Setmenge  $S \in \mathbb{S}$  definiert sein:

$$\forall S \in \mathbb{S} : |S| > 0$$

Weiterhin müssen sich die Eintrittswahrscheinlichkeiten aller *Sets* eines *Settypen* auf 1 summieren:

$$\forall S \in \mathbb{S} : \sum_{\sigma \in S} pr_{\sigma} = 1$$

Die für ein *Set*  $\sigma$  und einen Parameter  $p \in \mathbb{P}$  zugeordnete Annahmenverlaufsmenge  $A_{\sigma}^p$  enthält einzelne Annahmenverläufe  $a_{1,\sigma}^p, \dots, a_{v,\sigma}^p$  als Elemente. Ein Annahmenverlauf  $a_{j,\sigma}^p$  ist wiederum definiert als Tupel  $(pr_{j,\sigma}^p, V_{j,\sigma}^p)$ , wobei  $pr_{j,\sigma}^p \in (0, 1]$  wie schon bei der Definition eines *Sets* die Eintrittswahrscheinlichkeit des Annahmenverlaufs darstellt, wohingegen  $V_{j,\sigma}^p$  die angenommenen Werte enthält, also je nach Modellparameter  $p$  ein Element aus  $\mathbb{R}^t$  oder  $\mathbb{R}^{t \times x}$  ist.  $\mathbb{R}^t$  enthält dabei die angenommenen Werte eines Modellparameters in einem (eindimensionalen) Vektor über  $t$  Projektionsintervalle, wohingegen  $\mathbb{R}^{t \times x}$  die angenommenen Werte eines Modellparameters in einer (zweidimensionalen) Matrix über die Alter  $x$  und die Projektionsintervalle  $t$  enthält. Auch hier müssen die Eintrittswahrscheinlichkeiten auf 1 summieren:

$$\forall S \in \mathbb{S} \forall \sigma \in S \forall A_{\sigma}^p \in A_{\sigma} : \sum_{l=1}^{|A_{\sigma}^p|} pr_{l,\sigma}^p = 1$$

Um einen Prognosedurchlauf durchzuführen, muss nun für jeden *Settyp*  $P_i$  ein *Set*  $\sigma_i$  aus der entsprechenden Setmenge  $S_i$  ausgewählt werden. Die ausgewählten *Sets* seien mit  $\dot{\sigma}_i$  bezeichnet; alle für einen Prognosedurchlauf ausgewählten *Sets* seien im Tupel

$$\dot{S} = (\dot{\sigma}_1, \dots, \dot{\sigma}_k) \in S_1 \times \dots \times S_k \quad (10.35)$$

zusammengefasst. Die Eintrittswahrscheinlichkeit der ausgewählten Setkombination  $\dot{S}$  beträgt dann:

$$Pr[\dot{S}] = \prod_{\sigma_i \in \dot{S}} pr_{\sigma_i}. \quad (10.36)$$

Die Anzahl aller möglichen Setkombinationen entspricht dabei der in Gleichung 10.35 definierten Grundmenge:

$$N_{\mathbb{S}} = \prod_{i=1}^k |S_i|$$

Wenn eine gewählte Setkombination  $\dot{S}$  gegeben ist, kann die Menge der dafür auswählbaren Annahmenverlaufskombinationen ermittelt werden. Diese wird mit  $\mathbb{A}_{\dot{S}}$  bezeichnet und setzt sich wie folgt zusammen:

$$\mathbb{A}_{\dot{S}} = \prod_{i=1}^k \prod_{p_j \in P_i} A_{\dot{\sigma}_i}^{p_j} \quad (10.37)$$

wobei das jeweilige  $\dot{\sigma}_i$  das für den *Settyp*  $P_i$  gewählte *Set* darstellt, dessen Annahmenverlaufsmengen  $A_{\dot{\sigma}_i}^p$  ein Kreuzprodukt über alle in  $P_i$  zusammengefassten Modellparameter bilden. Diese *inneren* Kreuzprodukte in Gleichung 10.37 werden nun wiederum für alle *Settypen*  $P_1, \dots, P_k$  kombiniert, um die Menge aller möglichen Annahmenverlaufskombinationen zu erhalten, unter der Bedingung, dass die Setkombination  $\dot{S}$  gewählt wurde.

Im nächsten Schritt muss nun eine Annahmenverlaufskombination  $\dot{A}_{\dot{S}} = (\dot{a}^{q_1}, \dots, \dot{a}^{q_n}) \in \mathbb{A}_{\dot{S}}$  gewählt werden. Hierbei ist zu beachten, dass nun anstelle von  $p_1, \dots, p_n \in \mathbb{P}$  die Indizes  $q_1, \dots, q_n \in \mathbb{P}$  verwendet werden, da die Reihenfolge nicht identisch ist, d.h.  $p_i \neq q_i$  gelten kann (wegen Gleichung 10.37). Dies spielt für die kombinatorische Betrachtung jedoch keine Rolle.

Die Eintrittswahrscheinlichkeit von  $\dot{A}_{\dot{S}}$  kann nun, unter Berücksichtigung von Gleichung 10.36, wie folgt berechnet werden:

$$Pr[\dot{A}_{\dot{S}}] = Pr[\dot{S}] \cdot \prod_{a \in \dot{A}_{\dot{S}}} pr_a \quad (10.38)$$

Die Menge aller *insgesamt* möglichen Annahmenverläufe lässt sich nun durch die Kombination von Gleichung 10.35 und Gleichung 10.37 ermitteln:

$$N_{\mathbb{A}} = \sum_{\dot{S} \in S_1 \times \dots \times S_k} |\mathbb{A}_{\dot{S}}| \quad (10.39)$$

Der *Open Type* ist ein Spezialfall des *Limited Type*, bei dem es lediglich einen *Settyp*  $S'$  gibt, der alle Parameter beinhaltet ( $P' = \mathbb{P}$ ) und für den genau ein Set  $\sigma' = (1, A_{\sigma'})$  definiert ist.  $A_{\sigma'}$  enthält alle Annahmenverläufe für die Modellparameter. Da also für den *Open Type*  $|\mathbb{S}| = |\{S'\}| = 1$  und  $|S'| = 1$  gilt, gibt es nur einen Summand in Gleichung 10.39; demzufolge kann im *Open Type* die Anzahl aller möglichen Annahmenverlaufskombinationen folgendermaßen bestimmt werden:  $|\mathbb{A}_{\dot{S}}| = \prod_{p \in \mathbb{P}} |A_{\sigma'}^p|$  (vgl. Gleichung 10.37). So entspricht beispielsweise die Anzahl aller möglichen Annahmenverlaufskombinationen  $2^{246}$ , wenn bei einer Bevölkerungsprognose mit dem *Open Type* fünf Nachkommengenerationen (der direkten Zu- und Fortzüge) (und somit insgesamt 123 Modellparameter) unterschieden werden und

es für jeden Modellparameter 4 Annahmenverläufe gibt (d.h.  $\forall p \in \mathbb{P} : |A_{\sigma'}^p| = 4$ ) (vgl. Abschnitt 10.2.2):

$$N_{\mathbb{A}} = |\mathbb{A}_{S'}| = \prod_{p \in \mathbb{P}} |A_{\sigma'}^p| = \prod_{i=1}^{123} 4 = 4^{123} = (2^2)^{123} = 2^{246}$$

Wird in diesem Beispiel statt des *Open Type* jedoch der *Limited Type* verwendet, kann die Anzahl aller möglichen Annahmenverlaufskombinationen erheblich reduziert werden. Wenn nun beispielsweise für den alle Parameter umfassenden *Settyp*  $S'$  nicht nur ein Set  $\sigma'$ , sondern zwei *Sets*,  $\sigma'_1$  und  $\sigma'_2$ , definiert werden, die jeweils 2 Annahmenverläufe enthalten ( $|S'| = |\{\sigma'_1, \sigma'_2\}| = 2$ ), so gibt es trotz der insgesamt immer noch 4 Annahmenverläufe pro Modellparameter nur noch  $2^{124}$  statt  $2^{246}$  mögliche Annahmenverlaufskombinationen:

$$N_{\mathbb{A}} = |\mathbb{A}_{(\sigma'_1)}| + |\mathbb{A}_{(\sigma'_2)}| = \prod_{p \in \mathbb{P}} |A_{\sigma'_1}^p| + \prod_{p \in \mathbb{P}} |A_{\sigma'_2}^p| = \prod_{i=1}^{123} 2 + \prod_{i=1}^{123} 2 = 2 \cdot 2^{123} = 2^{124}$$

Damit ist in diesem Beispiel  $N_{\mathbb{A}}$  beim *Limited Type* um den Faktor  $2^{122}$  kleiner als beim *Open Type*. Doch trotz dieser Verbesserung bzw. Reduktion im *Limited Type* verbleiben insbesondere bei einer echten Bevölkerungsprognose (mit wesentlich mehr Annahmenverläufen pro Modellparameter) so viele mögliche Annahmenverlaufskombinationen, dass diese in der Regel nicht alle tatsächlich berechnet werden können.

### 10.3.2. Alternative Berechnungsmodi des Open Type und Limited Type

Wie im letzten Abschnitt illustriert, kann im PPPM die Anzahl potentieller Annahmenverlaufskombinationen durch die relativ hohe Anzahl der Modellparameter und der unbegrenzten Anzahl einzugebender Annahmenverläufe sehr groß sein.

Aus diesem Grund werden im PPPM die *simulative* Berechnung für heterogene Annahmenverteilungen (mit relativ vielen Annahmenverläufen pro Modellparameter) und die *vollständige bzw. erschöpfende* Berechnung für homogene Annahmenverteilungen (mit relativ wenigen Annahmenverläufen pro Modellparameter) alternativ angeboten. Darüber hinaus gibt es auch noch den Modus der *quasi-vollständigen* Berechnung zur erweiterten Ergebnisanalyse, bei dem ausschließlich die  $n$  Ergebnisverläufe mit der höchsten Eintrittswahrscheinlichkeit zu einer Ergebnisverteilung (für jede Ergebnisgröße) verbunden werden.

**Simulative Berechnung des PPPM**

Bei der simulativen Berechnung des PPPM erfolgt die Auswahl der Annahmenverläufe für jeden Modellparameter pro Prognosedurchlauf mit einer Monte-Carlo Simulation. Dazu wird vom Prognostiker im *Open Type* jedem Annahmenverlauf und im *Limited Type* zudem noch jedem Set eines Settyps eine Eintrittswahrscheinlichkeit subjektiv bzw. per Expertenmeinung zugeordnet. Generell liegt der Wert einer Eintrittswahrscheinlichkeit zwischen 0 und 1. Die Summe der Eintrittswahrscheinlichkeiten aller Annahmenverläufe eines Modellparameters muss 1 betragen genauso wie die Summe aller Eintrittswahrscheinlichkeiten aller Sets eines Settyps. Je höher der Wert der Eintrittswahrscheinlichkeit ist, desto höher schätzt ein Prognostiker das Eintreten des jeweiligen Annahmenverlaufs oder Sets ein. Gleichsam steigt die Wahrscheinlichkeit eines Annahmenverlaufs oder Sets, zufällig gezogen zu werden, mit der Höhe der jeweiligen Eintrittswahrscheinlichkeit. Auf diese Weise können in einer Bevölkerungsprognose mit dem PPPM sowohl sehr wahrscheinliche als auch weniger wahrscheinliche, aber dennoch potentielle, plausible und nicht ganz auszuschließende Annahmenverläufe subjektiv gewichtet berücksichtigt werden.

**Open Type** Die Auswahl eines Annahmenverlaufs erfolgt im Open Type des PPPM mit einem Zufallsexperiment für jeden Modellparameter pro Prognosedurchlauf. Bei jedem Zufallsexperiment wird eine Zufallszahl zwischen 0 und 1 nach einer Gleichverteilung mit einem Zufallszahlengenerator gezogen. Je höher die Eintrittswahrscheinlichkeit eines Annahmenverlaufs in Relation zu den anderen Annahmenverläufen eines zu betrachtenden Modellparameters ist, desto größer ist ihr Anteil am Bereich zwischen 0 und 1, aus dem die Zufallszahl jeweils zu ziehen ist; folglich ist ihre Chance bzw. Wahrscheinlichkeit, gezogen zu werden, größer als für andere Annahmenverläufe mit einer niedrigeren Eintrittswahrscheinlichkeit. Demnach bestimmen die Eintrittswahrscheinlichkeiten der Annahmenverläufe wie auch die jeweils zu ziehende Zufallszahl den jeweils auszuwählenden Annahmenverlauf für jeden Modellparameter pro Prognosedurchlauf. Zur Berechnung der Ergebnisgrößen des PPPM werden diese zufällig gezogenen Annahmenverläufe aller Modellparameter in das PPPM eingegeben. Diese Prozedur wird  $n$  Male wiederholt, um eine Ergebnisverteilung mit  $n$  Ergebnisverläufen pro Ergebnisgröße zu erhalten (vgl. Abbildung 3.3 auf Seite 44). Die Breite (bzw. Streuung) der resultierenden Ergebnisverteilung gibt dabei die Höhe der Prognoseunsicherheit einer Ergebnisgröße an.

Die Abbildung 10.3 zeigt exemplarisch die Auswahl eines Annahmenverlaufs zur Fertilität



für jede weibliche Subpopulation im *Open Type*. Dabei wird für jede weibliche Subpopulation ein Zufallsexperiment durchgeführt, bei dem jeweils ein Annahmenverlauf aus allen (für eine Subpopulation) einzugebenden Annahmenverläufen zur Fertilität (in Abhängigkeit der Eintrittswahrscheinlichkeiten) zufällig gezogen wird. Diese zufällig gezogenen Annahmenverläufe werden dann zur Berechnung eines Prognosedurchlaufs herangezogen. Das Auswahlverfahren findet in jedem Prognosedurchlauf für alle Modellparameter statt.

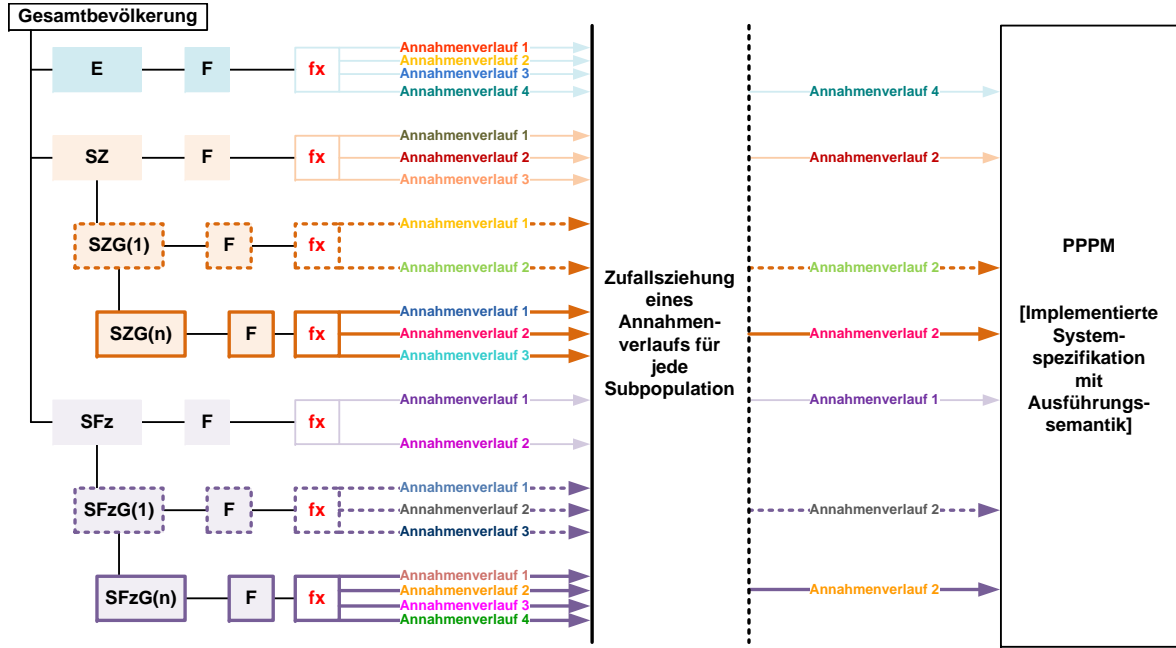


Abbildung 10.3.: Zufällige Auswahl eines Annahmenverlaufs im Open Type am Beispiel der *altersspezifischen Fertilitätsraten*: Für jede weibliche Subpopulation wird zufällig ein Annahmenverlauf zur Fertilität in Abhängigkeit von der subjektiv zugeordneten Eintrittswahrscheinlichkeit gezogen und zur Berechnung eines Ergebnisverlaufs für jede Ergebnisgröße in das PPPM (bzw. P3J) eingegeben. Für die einheimischen Frauen ist beispielsweise der petrolfarbene Annahmenverlauf 4 zufällig gezogen worden.

**Limited Type** Im *Limited Type* erfolgt die Auswahl eines Annahmenverlaufs für jeden Modellparameter über zwei Zufallsexperimente. In einem ersten Schritt muss für jeden Settyp ein Set in Abhängigkeit der jeweiligen Set-Eintrittswahrscheinlichkeiten gezogen werden. Dieses zufällig gezogene Set gilt dann für alle Modellparameter, die in dem jeweiligen Settyp zusammengeschlossen sind. In einem zweiten Schritt wird dann für jeden Modellparameter

eines Settyps ein Annahmenverlauf aus allen dem zufällig gezogenen Set zugeordneten Annahmenverläufen in Abhängigkeit ihrer Eintrittswahrscheinlichkeiten zufällig gezogen. Folglich gibt es nach diesem zweiten Zufallsexperiment für jeden Modellparameter (eines Sets eines Settypen) genau einen, zufällig gezogenen Annahmenverlauf, der zur Berechnung der Ergebnisgrößen in das Prognosemodell für einen Prognosedurchlauf eingegeben werden kann. Diese Prozedur wird in jedem Prognosedurchlauf für alle Settypen durchgeführt. Ist ein Modellparameter keinem Settyp zugeordnet, findet ein einfaches Zufallsexperiment zur Bestimmung eines Annahmenverlaufs pro Prognosedurchlauf statt. Generell werden auch beim *Limited Type*  $n$  Prognosedurchläufe mit immer wieder neuen Zufallsexperimenten durchgeführt, um für jede Ergebnisgröße eine Ergebnisverteilung mit  $n$  Ergebnisverläufen zu erhalten.

Die Abbildung 10.4 zeigt die Wahl eines Annahmenverlaufs zur Fertilität für jede weibliche Subpopulation im *Limited Type*. Zunächst werden in diesem Beispiel alle Modellparameter der Fertilität in einem Settyp *Fertilität* zusammengeschlossen. Anschließend werden die Annahmenverläufe der Modellparameter des Settyps Fertilität entweder dem pinken oder dem petrolfarbenen Set zugeordnet. Beispielsweise könnten Annahmenverläufe mit einer relativ hohen/niedrigen Fertilität im pinken/petrolfarbenen Set verbunden sein, wenn ein Prognostiker ähnlich hohe Fertilitätsniveaus in allen Subpopulationen erwartet. Mit dem ersten Zufallsexperiment wird nun für alle Modellparameter des Settyps *Fertilität* genau ein Set in Abhängigkeit der Set-Eintrittswahrscheinlichkeiten zufällig gezogen. In diesem Beispiel wird das petrolfarbene Set zufällig gezogen. Danach wird für jeden Modellparameter ein Zufallsexperiment durchgeführt, bei dem jeweils ein Annahmenverlauf aus allen (für diesen Modellparameter) einzugebenden Annahmenverläufen zur Fertilität aus dem petrolfarbenen Set (in Abhängigkeit der Eintrittswahrscheinlichkeiten) zufällig gezogen wird. Folglich wird für jeden Modellparameter des Settyps Fertilität ein petrolfarbener Annahmenverlauf zur Fertilität zufällig gezogen und zur Berechnung in das Prognosemodell eingegeben. Das Auswahlverfahren findet in jedem Prognosedurchlauf für alle in Settypen zusammengeschlossenen Modellparameter statt. Nach  $n$  Prognosedurchläufen entstehen  $n$  Ergebnisverläufe pro Ergebnisgröße.

**Ergebnisverteilung der Ergebnisgrößen** Ein Ergebnisverlauf resultiert aus der Kombination der zufällig gezogenen Annahmenverläufe pro Modellparameter. Je größer die Anzahl der einzugebenden Annahmenverläufe pro Modellparameter ist, desto größer ist

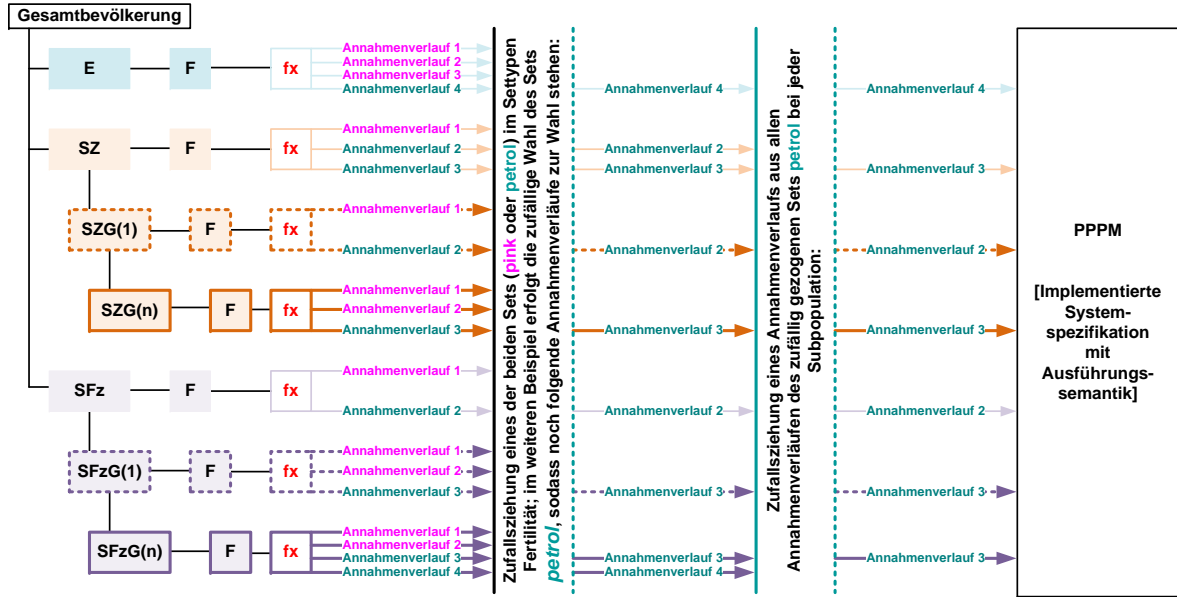


Abbildung 10.4.: Zufällige Auswahl eines Annahmenverlaufs im Limited Type am Beispiel der *altersspezifischen Fertilitätsraten*: In einem ersten Schritt wird für alle im Settyp Fertilität zusammengeschlossenen Subpopulationen zufällig ein Set (in Abhängigkeit von den Set-Eintrittswahrscheinlichkeiten) ausgewählt; in diesem Beispiel ist das petrolfarbene Set gezogen worden. In einem zweiten Schritt wird dann für jede weibliche Subpopulation zufällig ein Annahmenverlauf aus dem petrolfarbenen Set (in Abhängigkeit von den Annahmenverlaufs-Eintrittswahrscheinlichkeiten) ausgewählt, um mit dem PPPM (bzw. P3J) einen Ergebnisverlauf pro Ergebnisgröße zu berechnen; in diesem Beispiel ist der Annahmenverlauf 3 für die direkten Zuzüge, SZ, zufällig gezogen worden.

auch die Anzahl der zu unterscheidenden Ergebnisverläufe (vgl. Abschnitt 10.3.1). Um nun eine Ergebnisverteilung für jede Ergebnisgröße zu erhalten, werden  $n$  Prognosedurchläufe durchgeführt. Jeder Prognosedurchlauf erzeugt einen Ergebnisverlauf pro Ergebnisgröße, sodass  $n$  Prognosedurchläufe  $n$  Ergebnisverläufe generieren. Da bei den Prognosedurchläufen die Annahmenverläufe in Abhängigkeit ihrer Eintrittswahrscheinlichkeiten zufällig gezogen werden, entsteht die Mehrheit der Ergebnisverläufe einer Ergebnisgröße aus den Annahmenverläufen mit den höchsten Eintrittswahrscheinlichkeiten. Folglich werden die Annahmenverläufe mit einer geringeren Eintrittswahrscheinlichkeit seltener gezogen, sodass sie in weniger Ergebnisverläufen vertreten sind. Jeder Prognosedurchlauf kann unterschiedli-

che Ergebnisverläufe generieren, allerdings kann ein und derselbe Ergebnisverlauf auch rein zufällig mehrmals generiert werden. Um nun eine Ergebnisverteilung einer Ergebnisgröße abschätzen zu können, müssen nicht alle Ergebnisverläufe aus allen potentiellen Annahmenverlaufskombinationen berechnet werden. Stattdessen können gemäß dem *Gesetz der großen Zahlen* auch weniger Prognosedurchläufe durchgeführt werden, um die Ergebnisverteilung einer Ergebnisgröße simulativ zu schätzen. So nähert sich nach dem *Gesetz der großen Zahlen* die relative Häufigkeit eines Ergebnisverlaufs umso mehr seiner Wahrscheinlichkeit an, desto mehr Prognosedurchläufe durchgeführt werden. Die Anzahl der Prognosedurchläufe, die zur simulativen Abschätzung der Ergebnisverteilung pro Ergebnisgröße benötigt werden, hängt von der Anzahl der potentiellen Annahmenverlaufskombinationen wie auch von der Variabilität der Annahmenwerte und der Eintrittswahrscheinlichkeiten der einzugebenden Annahmenverläufe ab. Je höher die Anzahl der potentiellen Annahmenverlaufskombinationen und die Variabilität der einzugebenden Annahmenverläufe ist, desto höher ist die Anzahl der Prognosedurchläufe, die benötigt wird, um eine relativ konstante simulative Abschätzung der Ergebnisverteilung einer Ergebnisgröße zu erzielen. Eine simulativ zu schätzende Ergebnisverteilung einer Ergebnisgröße ist genau dann konstant bzw. sicher, wenn sie auch bei mehreren simulativen Abschätzungen mit dem PPPM immer wieder ohne größere Abweichungen herauskommt, da die relativen Häufigkeiten der Ergebnisverläufe ihren tatsächlichen Wahrscheinlichkeiten sehr nahe kommen.

Wenn für jede Ergebnisgröße eine Ergebnisverteilung simuliert worden ist, können so genannte Perzentil (Konfidenz) Intervalle bestimmt werden, die den Prognosewerten auch Eintrittswahrscheinlichkeiten zuordnen. Dazu werden die Werte der  $n$  Ergebnisverläufe einer Ergebnisgröße zu jedem Projektionsintervall unabhängig voneinander, der Größe nach aufsteigend sortiert. Um nun ein Perzentil (Konfidenz) Intervall  $[a, b]$  mit einer Irrtumswahrscheinlichkeit  $\alpha$  für eine Ergebnisgröße zu bestimmen, werden der  $(\frac{\alpha}{2})n$ -te und der  $(1 - \frac{\alpha}{2})n$ -te Wert der sortierten Ergebnisverläufe als unterer und oberer Grenzwert im jeweiligen Projektionsintervall genutzt. Somit liegen  $(1 - \alpha)100$  Prozent aller Ergebniswerte zwischen der oberen Grenze,  $a$ , und der unteren Grenze,  $b$ , des Perzentil (Konfidenz) Intervalls mit der Irrtumswahrscheinlichkeit  $\alpha$ . Die Eintrittswahrscheinlichkeit der Prognoseergebnisse innerhalb dieses Perzentil (Konfidenz) Intervalls entspricht dabei ihrer simulierten Häufigkeit.

**(Quasi-) Vollständige bzw. erschöpfende Berechnung des PPPM**

Bei der simulativen Berechnung des *Open Type* und des *Limited Type* werden zufällig  $n$  Ergebnisverläufe aus  $n$  Annahmenverlaufskombinationen berechnet. Dabei kann es vorkommen, dass eine Annahmenverlaufskombination mehrmals oder auch gar nicht gezogen und berechnet wird. Bei der (quasi-)vollständigen bzw. erschöpfenden Berechnung des *Open Type* und des *Limited Type* soll deshalb systematisch jede Annahmenverlaufskombination genau einmal mit ihrem exakten Gewicht berechnet werden und in die Ergebnisverteilung eingehen. Dabei sollen die Annahmenverlaufskombinationen in absteigender Reihenfolge hinsichtlich der Höhe ihrer Eintrittswahrscheinlichkeiten berechnet werden. Bei der vollständigen Berechnung wird jede Annahmenverlaufskombination berechnet, wohingegen bei der quasi-vollständigen Berechnung lediglich die  $n$  Annahmenverlaufskombinationen mit der höchsten Eintrittswahrscheinlichkeit berechnet werden. Die (quasi-)vollständige Berechnung ermöglicht im Gegensatz zur simulativen Berechnung die Generierung einer festen Ergebnisverteilung, die bei einer wiederholten Berechnung mit denselben Annahmen nicht von der vorherigen Ergebnisverteilung abweicht. Zudem wird bei der (quasi-)vollständigen Berechnung keine Rechnerleistung *vergeudet*, indem Ergebnisverläufe aus derselben Annahmenverlaufskombination mehrmals berechnet werden. Auf der anderen Seite ist die Umsetzung der (quasi-)vollständigen Berechnung für den *Open Type* und ganz besonders für den *Limited Type* sehr aufwendig, da der Algorithmus zur Suche der jeweils nächstwahrscheinlichen Annahmenverlaufskombination sehr komplex ist. Darüber hinaus stellt die verfügbare Rechnerleistung eine natürliche obere Grenze dar, die ab einer bestimmten Anzahl an Annahmenverlaufskombinationen erreicht ist, sodass ab da an nur noch die simulative und die quasi-vollständige Berechnung durchgeführt werden können.

**Open Type** Bei der (quasi-)vollständigen Berechnung des *Open Type* werden zunächst für jeden Modellparameter alle Annahmenverläufe in absteigender Reihenfolge hinsichtlich ihrer Eintrittswahrscheinlichkeit sortiert. Danach wird mit einem Breitensuchalgorithmus immer die Annahmenverlaufskombination mit der höchsten Eintrittswahrscheinlichkeit aus allen verbleibenden Annahmenverlaufskombinationen bestimmt. Dieser Breitensuchalgorithmus bestimmt als erstes die Annahmenverlaufskombination mit der höchsten bzw. maximalen Eintrittswahrscheinlichkeit. Dazu wird jeweils der Annahmenverlauf eines Modellparameters gewählt, der die höchste Eintrittswahrscheinlichkeit besitzt. Anschließend wird die Annahmenverlaufskombination mit der zweithöchsten Eintrittswahrscheinlichkeit gesucht. Dazu werden, gemäß der Breitensuche, die Eintrittswahrscheinlichkeiten der verbleibenden An-

annahmenverlaufskombinationen verglichen, die sich lediglich in einem Annahmenverlauf von der Annahmenverlaufskombination mit der höchsten Eintrittswahrscheinlichkeit unterscheiden. Aus diesen Annahmenverlaufskombinationen wird die Annahmenverlaufskombination mit der höchsten Eintrittswahrscheinlichkeit als zweithöchste bestimmt und berechnet, wohingegen die verbleibenden Annahmenverlaufskombinationen in eine Liste kommen, die die potentiellen Annahmenverlaufskombinationen mit den nächsthöchsten Eintrittswahrscheinlichkeiten enthält. Zu dieser Liste kommen bei der Suche der Annahmenverlaufskombination mit der dritthöchsten Eintrittswahrscheinlichkeit außerdem die Annahmenverlaufskombinationen, die sich in einem Annahmenverlauf von der Annahmenverlaufskombination mit der zweithöchsten Eintrittswahrscheinlichkeit unterscheiden. Diese Vorgehensweise wird solange durchgeführt bis die Liste mit den verbleibenden Annahmenverlaufskombinationen leer ist.

**Limited Type** Die (quasi-)vollständige Berechnung des *Limited Type* ist durch den Einbau der Settypen und Sets deutlich komplexer als beim *Open Type*. Hier werden die Annahmenverläufe der Modellparameter nicht nur nach ihrer eigenen Eintrittswahrscheinlichkeit, sondern auch nach der ihres Sets sortiert. Demzufolge können die in einer absteigenden Reihenfolge sortierten Annahmenverläufe pro Modellparameter auch aus verschiedenen Sets kommen. Im Anschluss an diese Sortierung werden die resultierenden Annahmenverlaufskombinationen von der höchsten bis zur niedrigsten Eintrittswahrscheinlichkeit nacheinander berechnet. Die Schwierigkeit besteht jedoch darin, dass im *Limited Type* ausschließlich *plausible* Annahmenverlaufskombinationen berechnet werden dürfen, um die Prognoseergebnisse nicht zu verzerren. In einer *plausiblen* Annahmenverlaufskombination können aber nur Annahmenverläufe desselben Sets von in einem Settyp zusammengeschlossenen Modellparametern miteinander kombiniert werden. So können die zu kombinierenden Sets bei einer plausiblen Annahmenverlaufskombination zwar zwischen Settypen, aber nicht innerhalb eines Settyps variieren. Folglich besteht die (quasi-)vollständige Berechnung des *Limited Type* nicht in der wahllosen Berechnung der Annahmenverlaufskombinationen mit der jeweils nächsthöchsten Eintrittswahrscheinlichkeit, sondern in der Berechnung der *plausiblen* Annahmenverlaufskombinationen mit der jeweils nächsthöchsten Eintrittswahrscheinlichkeit. Die Suche der *plausiblen* Annahmenverlaufskombination mit der jeweils nächsthöchsten Eintrittswahrscheinlichkeit entspricht dabei dem Prozedere im *Open Type*, nur dass unplausible Annahmenverlaufskombinationen verworfen werden.

**Ergebnisverteilung der Ergebnisgrößen** Lediglich die Erstellung der Ergebnisverteilung unterscheidet sich zwischen der simulativen und (quasi-)vollständigen Berechnung des PPPM, indem bei letzterer jede nur *einmalig* zu berechnende Annahmenverlaufskombination mit dem Gewicht ihrer Eintrittswahrscheinlichkeit in die Ergebnisverteilung (jeder Ergebnisgröße) eingeht. Ansonsten weicht die Aufbereitung bzw. Darstellung der (quasi-)vollständig zu berechnenden Ergebnisverteilung nicht von der simulativ zu berechnenden Ergebnisverteilung ab, sodass auch hier beispielsweise verschiedene Perzentil (Konfidenz) Intervalle bestimmt werden können.

## 10.4. P3J — Berechnung der Prognoseergebnisse

Wie bereits in Abschnitt 9.3.1 beschrieben, setzt das P3J auf dem Modellierungs- und Simulationsrahmenwerk JAMES II auf. Die sich daraus ergebenden Vorteile werden besonders bei der Entwicklung der Algorithmen zur Berechnung der Ergebnisverteilungen (für die wesentlichen Ergebnisgrößen) sichtbar.

Das P3J soll den simulativen wie auch den (quasi-)vollständigen Berechnungsmodus zur Bestimmung einer Ergebnisverteilung für die wesentlichen Ergebnisgrößen unterstützen (vgl. Abschnitt 10.3.2). Dazu wird für jeden Berechnungsmodus ein Algorithmus entwickelt, der sowohl für den *Open Type* als auch für den *Limited Type* angewendet werden kann. Der *Open Type* wird dabei als ein Spezialfall des *Limited Type* mit nur einem Settyp und einem Set betrachtet (vgl. Abschnitt 10.3.1). Über die Bedienoberfläche soll der Nutzer letztlich entscheiden, ob die Prognoseergebnisse mit dem simulativen oder (quasi-)vollständigen Ansatz berechnet werden und wie viele Prognosedurchläufe durchgeführt werden sollen.

Mit diesen Einstellungen des Nutzers wird ein *allgemeiner* Algorithmus zur Berechnung der Ergebnisverläufe (wesentlicher Ergebnisgrößen) initialisiert. Dieser Algorithmus führt folgende Arbeitsschritte sukzessive (bei jedem Berechnungsmodus) aus:

1. Wähle einen Annahmenverlauf pro Modellparameter
2. Berechne den Ergebnisverlauf für die gewählte Annahmenverlaufskombination (vgl. Abschnitt 10.1)
3. Speichere die Prognoseergebnisse

4. Falls die Anzahl der bereits durchgeführten die der insgesamt durchzuführenden Prognosedurchläufe unterschreitet, beginne von vorn

Dieser allgemeine Algorithmus ist als Plug-in, also als wiederverwendbare Software-Komponente, zur Berechnung von PPPM-Prognosen in JAMES II neu definiert (vgl. Abschnitt 9.3.1). Damit JAMES II auch die Eingabedaten (vgl. Abschnitt 9.3.2) als PPPM-Daten erkennt und an den oben beschriebenen Algorithmus weitergibt, muss ein zweites Plug-in zum Einlesen von (Prognose-)Modellen erstellt werden. Durch die Definition dieser zwei neuen JAMES II Plug-ins können nun viele der im Rahmenwerk vordefinierten Arbeitsabläufe auf die PPPM-Eingabedaten und deren Berechnung angewendet werden. So werden die JAMES II Komponenten zum Experimentieren verwendet, um die Berechnung gegebenenfalls auf mehrere Prozessorkerne auszulagern (und durch eine parallelisierte Bearbeitung zu beschleunigen). Die Experimentierkomponenten sind auch für die Steuerung der Berechnung, also deren Pausieren, Abbruch, oder Beendigung, zuständig. Alle diese Funktionen werden automatisch von JAMES II bereitgestellt und müssen nicht neu entwickelt werden.

Das Plug-in System von JAMES II wird außerdem verwendet, um den allgemeinen Algorithmus mit den alternativen Berechnungsmodi zur Auswahl einer Annahmenverlaufskombination zu assoziieren. Dazu werden der simulative und (quasi-)vollständige Ansatz ebenfalls in einem JAMES II Plug-in realisiert und dem allgemeinen Algorithmus als entsprechende Hilfskomponente zugewiesen. Die Abbildung 10.5 zeigt dazu wie die (hier ausgewählten) JAMES II Komponenten um die extra zu realisierenden P3J-Komponenten ergänzt werden, sodass überhaupt eine Bevölkerungsprognose (im Sinne des PPPM) ausgeführt werden kann. Zudem wird auch der Einschluss externer Komponenten wie z. B. Hibernate (vgl. Abschnitt 9.3.2) dargestellt.

Die nächsten Abschnitte erläutern die technische Umsetzung der in Abbildung 10.5 skizzierten P3J-Komponenten zur Berechnung einer Ergebnisverteilung wie auch die grundlegende Funktionsweise wichtiger Hilfskomponenten (z. B. zur Erzeugung von (Pseudo-)Zufallszahlen) für den simulativen und (quasi-)vollständigen Berechnungsmodus.

### 10.4.1. Simulativer Ansatz

Um den simulativen Ansatz zur Berechnung einer Ergebnisverteilung umzusetzen, muss eine *zufällige* Kombination aus Annahmenverläufen gezogen werden, wobei jeder zufällig



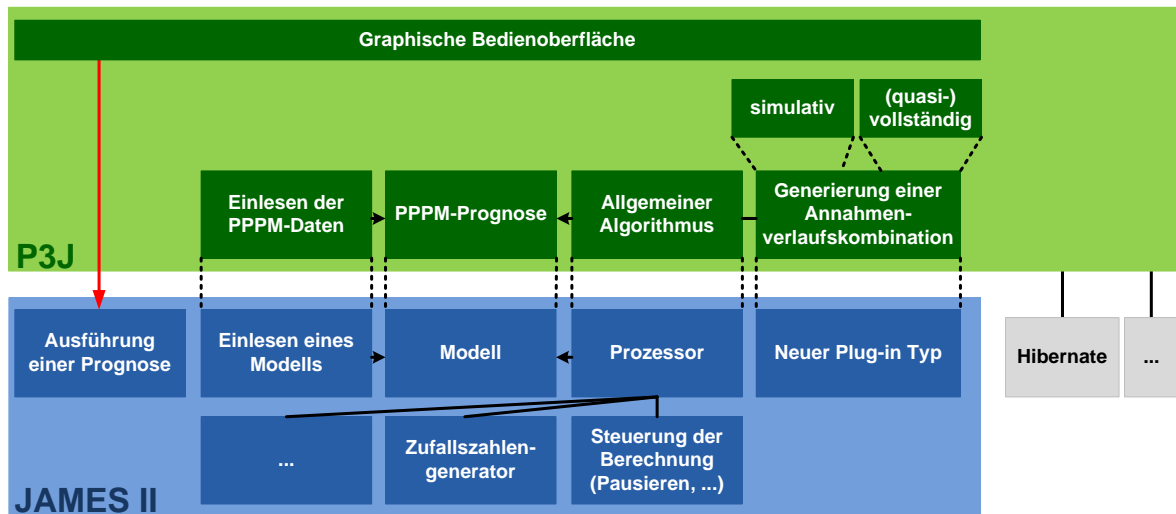


Abbildung 10.5.: Realisierung der P3J-Funktionen mit JAMES II: Definition neuer Plug-ins zum Einlesen von Eingabedaten (in das Modell) und zur Berechnung der Ergebnisverläufe (mit dem Prozessor), um die Experimentierkomponenten von JAMES II (z. B. zur Steuerung der Berechnung oder auch zur Generierung von Zufallszahlen) verwenden zu können. Diese werden aufgerufen, sobald der Nutzer die Berechnung aus der P3J-Bedienungsumgebung heraus startet (roter Pfeil). Das Plug-in System von JAMES II ist außerdem verwendet worden, um die alternativen Modi zur Auswahl von Annahmenverlaufskombinationen mit dem allgemeinen Berechnungsansatz zu verbinden.

zu ziehende Annahmenverlauf mit einer Eintrittswahrscheinlichkeit gewichtet ist; eine so getroffene Auswahl entspricht einem klassischen Monte-Carlo Ansatz, da zufällig ein Element der Menge  $N_A$  (vgl. Gleichung 10.39) gezogen wird (vgl. Abschnitt 10.3.2 ab Seite 208). Die Größe und Struktur von  $N_A$  hängt dabei von den definierten Settypen und Sets ab (vgl. Abschnitt 10.3.1). Generell kann die zufällige Auswahl als zweistufiger Prozess angesehen werden, bei dem zunächst für jeden Settyp ein Set und danach für jedes Set eine Annahmenkombination (für die im entsprechenden Settyp eingeschlossenen Modellparameter) gezogen wird, wobei der erste Schritt im Fall des *Open Type* entfällt (vgl. Abschnitt 10.3.2). Zur Umsetzung dieses *zufälligen* Auswahlprozesses wird eine zusätzliche Hilfskomponente benötigt, die so genannte Pseudo-Zufallszahlen generiert.

Wie der Name bereits andeutet, erzeugt ein (Pseudo-)Zufallszahlengenerator Sequenzen von Zahlen, die lediglich zufällig *erscheinen*. Digitalcomputer sind ohne spezielle Zusatzkomponenten nicht in der Lage, wirklich zufällige Zahlenfolgen zu generieren, da sie als deterministische Zustandsmaschinen konzipiert sind.<sup>8</sup> Generatoren von Pseudo-Zufallszahlen<sup>9</sup> begegnen dieser Beschränkung, indem sie Funktionen zur Berechnung einer Zahlenfolge verwenden, die diese zufällig erscheinen lassen. Gewöhnlich werden Fließkommazahlen, die gleichverteilt im Intervall  $[0, 1)$  liegen, erzeugt. Diese werden dann mit weiteren Methoden kombiniert, um Zufallszahlenfolgen mit anderen Charakteristika zu erzeugen.

Die Sequenz der generierten Zahlen hängt vom Startzustand des Generators, dem so genannten *Seed*, ab. Soll eine Zahl generiert werden, wird zunächst eine Funktion zum Wechseln vom aktuellen Zustand des Generators in einen neuen Zustand ausgeführt. Dieser neue Zustand (oder Teile davon) werden dann zur Berechnung einer weiteren Zahl in der Folge genutzt. Da ein Zustandsautomat nur endlich viele Zustände besitzen kann, wird nach einer gewissen (normalerweise sehr großen) Anzahl von Zustandsänderungen, d. h. nach einer großen Anzahl von generierten Zahlen, ein vorher bereits eingenommener Zustand erreicht. Wenn dies geschieht, wiederholt sich die Sequenz an Zufallszahlen von diesem Punkt an, da der Algorithmus an sich deterministisch ist. Die Anzahl der Zustandsänderungen bis zu solch einer Wiederholung wird deshalb auch als Periodenlänge bezeichnet. Für die praktische Anwendung der in dieser Arbeit behandelten Ansätze ist die Periodenlänge aktueller Zufallszahlengeneratoren ausreichend.<sup>10</sup> Sind der verwendete Generator und der Seed bekannt, lässt sich die vom Generator erzeugte Sequenz an (Pseudo-)Zufallszahlen eindeutig reproduzieren. Diese Eigenschaft bringt viele Vorteile bei der Entwicklung und dem Testen von Programmen mit sich, sodass die Generierung von *echten* (d. h. nicht reproduzierbaren) Zufallszahlen nur noch in wenigen Anwendungsfällen notwendig ist.

Obwohl deterministische Zufallszahlengeneratoren viele Vorteile bieten, sollten sie umsichtig verwendet werden. Neben der bereits erwähnten Periodenlänge, die nicht überschritten werden darf,<sup>11</sup> sollte auch die Qualität der generierten Zahlen überprüft werden. Es muss

---

<sup>8</sup>Im Gegensatz zu Analogcomputern, bei denen z. B. das thermische Rauschen eines Widerstands als Quelle für Zufallszahlen verwendet werden kann ([141], S. 3).

<sup>9</sup>Vereinfachend werden die Generatoren von Pseudo-Zufallszahlen nachfolgend als Zufallszahlengeneratoren bezeichnet.

<sup>10</sup>Sie ist meist größer als  $2^{31}$  [84].

<sup>11</sup>Genauer gesagt sollte die Anzahl der durch einen Generator mit der Periodenlänge  $n$  gezogenen Zufallszahlen sehr viel kleiner als  $n$  sein (bei einigen Generatoren z. B. nicht größer als  $\sqrt{n}$  ([106], S. 495)).

sichergestellt werden, dass diese *zufällig* genug sind, um einen systematischen Fehler bei der Ermittlung der Stichproben zu vermeiden. Die Vermeidung solcher Korrelationen hängt stark von der Art und Weise ab, wie der Generator bzw. dessen Ausgabe genutzt wird [106]; im konkreten Fall also z. B. in Verbindung mit den Eintrittswahrscheinlichkeiten der Annahmenverläufe und Sets, die bei der simulativen Berechnung zufällig ausgewählt werden.

Da die Eignung eines Zufallszahlengenerators vom jeweiligen Kontext abhängt, sollte nicht ein Generator a priori ausgewählt werden. Anstatt dessen ist es empfehlenswert, *mehrere* Zufallszahlengeneratoren unterschiedlichen Typs zu verwenden und deren Ergebnisse zu vergleichen ([106], S. 502). Dies erlaubt bei Bedarf die Überprüfung etwaiger Unstimmigkeiten in den Ergebnissen. Gleichzeitig sollte die korrekte Programmierung der verwendeten Zufallszahlengeneratoren durch statistische Tests ([141], S. 41) sichergestellt werden. Werden die genannten Aspekte vernachlässigt, besteht die Gefahr, dass die generierte Folge an Zufallszahlen einen unbeabsichtigten (und unerkannten) Trend enthält, der die Ergebnisse der Berechnungen verzerrt bzw. verfälscht ([100]).

JAMES II bietet eine ganze Reihe an ausführlich getesteten Zufallszahlengeneratoren [84]. Für den simulativen Ansatz wird als Standardeinstellung der JAMES II Standard-Generator, der dem Standard-Generator der Java-Klassenbibliothek entspricht, verwendet.<sup>12</sup>

#### 10.4.2. (Quasi-)Vollständiger Ansatz

Alternativ zum simulativen Ansatz können die Annahmenverläufe auch nach einem bestimmten *Muster* bzw. nach einer bestimmten Reihenfolge ausgewählt werden. So kann es für eine weitere Analyse der Prognoseergebnisse z. B. zweckmäßig sein, die Elemente von  $N_{\mathbb{A}}$  in absteigender Reihenfolge ihrer Eintrittswahrscheinlichkeiten auszuwählen. Die grundlegende Vorgehensweise bei dieser vollständigen Berechnung des PPPM, die jedoch häufig aufgrund ihres Umfangs nach einer festzulegenden Anzahl an Prognosedurchläufen (in der dann *quasi*-vollständigen Variante) abubrechen ist, ist bereits in Abschnitt 10.3.2 beschrieben worden.

Für die Umsetzung des (quasi-)vollständigen Ansatzes wird der Open Type wiederum als ein Spezialfall des Limited Type mit lediglich einem Settyp und Set angesehen. Der Algorith-

---

<sup>12</sup>Dabei handelt es sich um einen *linearen Kongruenzgenerator* ([141, S. 10-11]).

mus für die (quasi-)vollständige Berechnung des Limited Type ist zweistufig. Auf der oberen Ebene werden alle Setkombinationen bzw. die Elemente  $\dot{S}$  in absteigender Reihenfolge ihrer Eintrittswahrscheinlichkeiten sortiert (vgl. Gleichungen 10.35 und 10.36). Auf der unteren Ebene wird dann für jedes Set  $\sigma_i$  eines Settyps  $S_j$  die Reihenfolge der wahrscheinlichsten Annahmenverlaufskombinationen ermittelt; eine Annahmenverlaufskombination wird dabei aus den Annahmenverlaufsmengen  $A_{\sigma_i}^p$  für alle Parameter  $p \in P_j$  aus  $\sigma_i$  gebildet. Die Eintrittswahrscheinlichkeit einer solchen Annahmenverlaufskombination  $\dot{a}_{\sigma_i} = (\dot{a}^{p_{i_1}}, \dots, \dot{a}^{p_{i_n}})$ , mit  $n = |P_j|$  und  $i_1, \dots, i_n$  als Indizes der Parameter in  $P_j$ , ergibt sich aus:

$$\prod_{j=1}^n pr_{\dot{a}^{p_{i_j}}}, \dot{a}_{i_j}^p \in A_{\sigma_i}^{p_j}$$

Die endgültige Eintrittswahrscheinlichkeit einer Annahmenverlaufskombination wird abschließend aus dem Produkt der Eintrittswahrscheinlichkeit der ausgewählten Sets  $\dot{S}$  (vgl. Gleichung 10.36) und der ausgewählten Annahmenverläufe von jedem ausgewählten Set  $\dot{\sigma} \in \dot{S}$  (vgl. Gleichung 10.35 und 10.38) berechnet. Um sicherzustellen, dass die Reihenfolge, die sich aus diesen endgültigen Eintrittswahrscheinlichkeiten ergibt, richtig berechnet wird, werden auf beiden Ebenen die Mengen an bereits generierten Set- bzw. Annahmenverlaufskombinationen zwischengespeichert. Dies ist notwendig, da Annahmenverlaufskombinationen mit kleineren Eintrittswahrscheinlichkeiten auch Teile von wahrscheinlicheren, bereits ermittelten Annahmenverlaufskombinationen enthalten. Durch das Zwischenspeichern der bereits bekannten Teilkombinationen kann somit Rechenzeit gespart werden. Neben der langen Laufzeit, die trotz des Zwischenspeicherns aufgrund der Vielzahl der zu berechnenden Annahmenverlaufskombinationen hervorgerufen wird, ist auch der Speicherbedarf dieses Algorithmus relativ hoch. Dies liegt am Zwischenspeichern der Teilkombinationen, anhand derer sich die *insgesamt* höchste bzw. wahrscheinlichste unberechnete Annahmenverlaufskombination ermitteln lässt.

Um diese Beschränkungen zu umgehen, wird eine *quasi*-vollständige Analyse ermöglicht. Hierbei wird nach einer bestimmten Anzahl von Prognosedurchläufen (bzw. einer bestimmten kumulierten Eintrittswahrscheinlichkeit aller betrachteten Annahmenverlaufskombinationen) automatisch abgebrochen (vgl. Abschnitt 10.3.2 ab Seite 213). Im Gegensatz zur simulativen und vollständigen Berechnung sollte die quasi-vollständige Berechnung nicht zur Bestimmung einer *kompletten* Ergebnisverteilung, sondern vorrangig zur erweiterten Ergebnisanalyse verwendet werden (vgl. Abschnitt 10.3.2 ab Seite 213).

### 10.4.3. Überprüfung der Eintrittswahrscheinlichkeiten

Um die Verwendung des P3J zu vereinfachen, wird sowohl beim simulativen als auch beim (quasi-)vollständigen Ansatz zuvor überprüft, ob die eingegebenen Eintrittswahrscheinlichkeiten für die Annahmenverläufe (eines Modellparameters) und Sets (eines Settyps) auf 1 summieren (vgl. Abschnitt 10.3.1). Ist dies nicht der Fall, sind zwei Situationen zu unterscheiden. Wenn die Summe der Eintrittswahrscheinlichkeiten,  $s_{pr}$ , größer als 0 ist, werden sie entsprechend der eingegebenen Gewichtung auf 1 normalisiert, d. h. jede Eintrittswahrscheinlichkeit wird mit dem Faktor  $\frac{1}{s_{pr}}$  multipliziert. Wenn  $s_{pr}$  hingegen gleich 0 ist, werden stattdessen gleichverteilte Eintrittswahrscheinlichkeiten angenommen. So erhält beispielsweise jeder der  $n$  Annahmenverläufe eines Modellparameters die Eintrittswahrscheinlichkeit  $\frac{1}{n}$ . In beiden Fällen wird der Nutzer durch eine Warnmeldung über die vorgenommene Skalierung informiert.

### 10.4.4. Unterstützung von repräsentativen Annahmenverläufen

Die in Abschnitt 10.2.1 beschriebenen repräsentativen Annahmenverläufe, die eine zufällige proportionale Variation der eingegebenen (, repräsentativen) Annahmenwerte erlauben, müssen ebenfalls von den Berechnungskomponenten unterstützt werden. Hierbei wird für jeden Annahmenverlauf die vom Nutzer festgelegte Standardabweichung berücksichtigt. Beträgt die Standardabweichung null, werden die eingegebenen Annahmenwerte für jeden Prognosedurchlauf so übernommen, wie sie eingegeben worden sind. Andernfalls werden die eingegebenen (, repräsentativen) Annahmenwerte durch das Ziehen von Zufallszahlen (vgl. Abschnitt 10.4.1) in dem vom Nutzer gewünschten Ausmaß variiert. Dies geschieht, indem die repräsentativen Annahmenwerte mit einem Fehlervektor  $\vec{x}$  multipliziert werden, der von einem einfachen stochastischen Prozess generiert wird. Der Prozess startet im Zustand  $x_0 = 1$ . Jeder weitere Zustand  $x_i$  wird berechnet mit:

$$x_i = x_{i-1} \cdot \max(\epsilon, X), X \sim N(1, \sigma), 1 \gg \epsilon > 0$$

wobei die Standardabweichung der Normalverteilung, also  $\sigma$ , dem vom Nutzer einzugebenden Maß an Variation entspricht. Die Variable  $\epsilon > 0$  sichert zu, dass  $x_i \neq 0$  gilt. Bezieht sich der Annahmenverlauf auf die *Mortalität* einer Subpopulation, werden die repräsentativen Annahmenwerte zunächst von den Überlebenden im Alter  $x$  in logarithmische Sterbewahrscheinlichkeiten umgewandelt, die dann anschließend mit den Elementen des Fehlervektors multipliziert werden. Danach wird die Konvertierung rückgängig gemacht. Während die Un-

terstützung von repräsentativen Annahmenverläufen nur wenig Einfluss auf den simulativen Ansatz zur Berechnung (vgl. Abschnitt 10.4.1) hat, da höchstens die Anzahl der für aussagekräftige Ergebnisse benötigten Prognosedurchläufe steigt, ist sie für den (quasi-)vollständigen Ansatz (vgl. Abschnitt 10.4.2) problematisch. Das liegt daran, dass der (quasi-)vollständige Ansatz jede Kombination von Annahmenverläufen höchstens *einmal* berechnet. Zwar kommt ein Annahmenverlauf meist in einer Vielzahl von Annahmenverlaufskombinationen vor, so dass trotzdem mehrere zufällige Variationen der repräsentativen Annahmenwerte generiert werden, aber dies gilt nicht immer. Eine alternative Herangehensweise wäre deshalb, die Variation eines repräsentativen Annahmenverlaufs beim (quasi-)vollständigen Ansatz völlig außen vor zu lassen.

## 11. Prognoseergebnisse des PPPM

Nachdem in den Kapiteln 9 und 10 die Erzeugung der Prognoseergebnisse (über die Annahmenverlaufskombinationen der Modellparameter) im Fokus gestanden hat, befasst sich das Kapitel 11 mit der Verwendung der Prognoseergebnisse bzw. den Möglichkeiten zur Ergebnispräsentation und erweiterten Ergebnisanalyse.

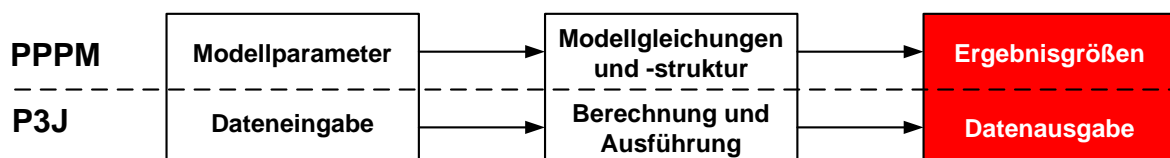


Abbildung 11.1.: Aspekte des theoretischen Modells, PPPM, und seiner Realisierung P3J:  
Das aktuelle Kapitel 11 befasst sich mit den Prognoseergebnissen des PPPM und P3J (rot unterlegt).

Bei einer probabilistischen Bevölkerungsprognose mit dem PPPM sind die Prozesse der Annahmengenerierung, der Berechnung und Speicherung bzw. Verwaltung der Prognoseergebnisse weitestgehend voneinander getrennt, um jeweils eine flexible und uneingeschränkte Ausführung zu ermöglichen.

Das PPPM bietet eine detaillierte Ergebnisanalyse, da nicht nur die Annahmenverläufe (vgl. Abschnitt 9.3.2 ab Seite 179), sondern auch die Prognossergebnisse (aller Ergebnisverläufe) in einer externen Datenbank gespeichert werden. Dabei werden für jeden Ergebnisverlauf sowohl die prognostizierten Werte wesentlicher Ergebnisgrößen (vgl. Abschnitt 11.1) als auch die jeweils gezogenen Annahmenverläufe pro Modellparameter gespeichert. Dadurch kann (wie in herkömmlichen Bevölkerungsprognosemodellen auf der Makroebene) die zukünftige Entwicklung der Gesamtbevölkerung, aber auch zusätzlich die einzelner (und aggregierter) Subpopulationen nach Alter und Geschlecht (im Längsschnitt wie auch im Querschnitt) analysiert werden; ein optional vom PPPM zu erstellender Ergebnisreport fasst diese grundlegenden Prognoseergebnisse automatisch in einem Dokument zusammen (vgl. Abschnitt 11.2).

Darüber hinaus können mit dem PPPM (im Gegensatz zu herkömmlichen Bevölkerungsprognosemodellen) auch Ergebnisverläufe extrahiert und zu so genannten *Miniprojektionen* zusammengeschlossen werden, die alle einen oder mehrere bestimmte Annahmenverläufe eines oder mehrerer Modellparameter gemein haben. Außerdem kann z.B. mit einem Levelplot analysiert werden, welche Annahmenkombinationen niedrige, mittlere und hohe Prognosewerte beliebiger Ergebnisgrößen erzeugen. Somit können bei einer Bevölkerungsprognose mit dem PPPM nicht nur die Prognoseergebnisse separat, sondern auch variabel aggregiert und zusammen mit den sie erzeugenden Annahmen genau analysiert werden.

### 11.1. Ergebnisgrößen

Im PPPM werden die Jahresend- und Jahresdurchschnittsbestände nach Alter und Geschlecht für die einzeln zu prognostizierenden Subpopulationen, also für die Einheimischen, die direkten Zu- und Fortzüge wie auch für deren einzelne Nachkommengenerationen als wesentliche Ergebnisgrößen (vgl. Abschnitt 10.1) für jeden Prognosedurchlauf extern in einer Datenbank gespeichert.

Ausgehend von diesen wesentlichen Ergebnisgrößen können die Jahresend- und Jahresdurchschnittsbestände weiterer (Sub)Populationen variabel aggregiert bzw. berechnet und analysiert werden. So können z.B. die Gesamtbevölkerung, alle Zu- bzw. Fortzüge, alle Nachkommen(generationen) der direkten Zu- bzw. Fortzüge wie auch die Salden dieser einzelnen und aggregierten Zu- und Fortzugssubpopulationen gebildet werden (vgl. Tabelle 11.1).

Folglich kann mit dem PPPM nicht nur (wie bei herkömmlichen Bevölkerungsprognosemodellen) die zukünftige Entwicklung der Gesamtbevölkerung, der Nettomigration und der Einheimischen, sondern auch das tatsächliche Aufkommen aller Immigranten und Emigranten und deren Zusammensetzung aus direkten Zu- und Fortzügen und deren einzelnen Nachkommengenerationen genau prognostiziert werden.

Da für jede der einzeln prognostizierten und variabel aggregierten (Sub)Populationen die Bestände nach Einzelalter und Geschlecht vorliegen, können daraus auch noch weitere Ergebnisgrößen, wie z.B. diverse Alterungsmaße, berechnet werden. So kann beispielsweise das *Old Age Dependency Ratio* aus dem Verhältnis der 65+-Jährigen zu den 15 – 64-Jährigen berechnet werden, um das Verhältnis der zu versorgenden zur versorgenden Bevölkerung



	(Sub)Populationen
separat prognostiziert	$E$
	$SZ$
	$SFz$
	$SZG(n)$
	$SFzG(n)$
variabel aggregiert	$P_{gesamt}$
	$\sum_{n=1}^a SZG(n)$
	$\sum_{n=1}^a SFzG(n)$
	$\sum_{n=1}^a SZG(n) + SZ$
	$\sum_{n=1}^a SFzG(n) + SFz$
	$SZ - Fz$
	$SZG(n) - SFzG(n)$
	$\sum_{n=1}^a SZG(n) - \sum_{n=1}^a SFzG(n)$
	$\sum_{n=1}^a SZG(n) + SZ - \sum_{n=1}^a SFzG(n) - SFz$

Tabelle 11.1.: Prognostizierte Jahresend- und Jahresdurchschnittsbestände für die einzeln prognostizierten und daraus variabel aggregierten (Sub)Populationen;  $P_{gesamt}$  entspricht der Gesamtpopulation und  $a$  entspricht jeweils der höchsten Nachkommengeneration der direkten Zu- bzw. Fortzüge.

für einzelne (Sub)Populationen im Zeitverlauf oder aber auch im direkten Vergleich zu analysieren. Anhand geeigneter demographischer Alterungsmaße können demnach potentielle zukünftige Umwandlungsprozesse wie die demographische Alterung erkannt und quantifiziert werden, sodass ausgleichende politische Handlungen vorausschauend geplant und umgesetzt werden können. In diesem Zusammenhang ist z.B. auch eine Prognose der zukünftigen Schulanfänger, Erwerbspersonen oder auch Pflegebedürftigen zu erwähnen, die als Anteil bestimmter Alters- und Geschlechtsgruppen in den jeweiligen (Sub)Populationen bestimmt werden können.

Darüber hinaus kann z.B. der bevölkerungsdynamische Einfluss der (Sub)Populationen auf die Entwicklung der Gesamtbevölkerung ermittelt werden; so kann beispielsweise untersucht werden, inwiefern die (durch eine sehr niedrige Fertilität verursachte) Schrumpfung einer einheimischen Bevölkerung von Immigranten (mit einer höheren Fertilität für die direkten Zuzüge und deren erste und zweite Nachkommengeneration) gedämpft bzw. abgefangen werden kann. Darüber hinaus kann z.B. auch der bevölkerungsdynamische Effekt der

Annahmen zur Sterblichkeit, Fertilität und Migration auf die zukünftige Entwicklung einer (Sub)Population ermittelt werden: Um beispielsweise den bevölkerungsdynamischen Effekt der Sterblichkeitsannahmen zu untersuchen, wird die prognostizierte Entwicklung infolge der Annahmen zur Sterblichkeit, Fertilität und Migration mit der prognostizierten Entwicklung verglichen, bei der die Sterblichkeit aus dem Jump-off Jahr konstant beibehalten wird und lediglich die Annahmen zur Fertilität und Migration berücksichtigt werden. Da ausschließlich die Sterblichkeit die Differenz zwischen diesen beiden prognostizierten Entwicklungen bewirkt, wird dieser Effekt auch als Sterblichkeitseffekt bezeichnet.

### 11.2. Ergebnisreport

Zur Reduktion des Arbeitsaufwands kann im PPPM bzw. P3J optional für jede Bevölkerungsprognose ein Ergebnisreport mit häufig zu verwendenden Abbildungen wesentlicher Ergebnisgrößen (vgl. Abschnitt 11.1) automatisch als *pdf*-Dokument erstellt werden. Damit wird die Ergebnisanalyse insofern erleichtert, als dass häufig benötigte Prognoseergebnisse zeitsparend automatisch erzeugt werden, ohne diese immer wieder zeitintensiv manuell mit diversen Softwarepaketen zur statistischen Analyse erstellen zu müssen.

Der Ergebnisreport fasst die prognostizierte Entwicklung der Jahresendbestände der einzelnen zu prognostizierenden und daraus variabel zu aggregierenden (Sub)Populationen aus der Tabelle 11.1 graphisch zusammen. Für jede dieser Ergebnisgrößen werden die prognostizierten Werte ausgewählter Quantile ( $p = 0, 0,025, 0,05, 0,01, 0,5, 0,9, 0,95, 0,975, 1$ ) ausgegeben, die dem Median wie auch den häufig verwendeten 80–, 90–, 95– und 100– prozentigen Perzentil (Konfidenz) Intervallen entsprechen. Ein x-prozentiges Perzentil (Konfidenz) Intervall beinhaltet dabei x Prozent aller Ergebnisverläufe um den jeweiligen Median einer Ergebnisverteilung (vgl. Abschnitt 10.3.2 ab Seite 210).

Sollten neben den im Ergebnisreport aufgeführten noch weitere Prognoseergebnisse benötigt werden, ist dafür auf die in der Datenbank gespeicherten Daten zurückzugreifen.

Im Ergebnisreport werden für jede (Sub)Population die prognostizierten Jahresendbestände in drei Abbildungen (mit den genauen Prognoseergebnissen ausgewählter Quantile und Projektionsintervalle) dargestellt.

Nachfolgend sollen diese verschiedenen Abbildungen des Ergebnisreports kurz anhand der Prognoseergebnisse der Variante II der Beispielprojektion aus dem Kapitel 12 erläutert werden.<sup>1</sup>

### 11.2.1. Entwicklung der Ergebnisverteilung diverser Ergebnisgrößen über den Projektionshorizont

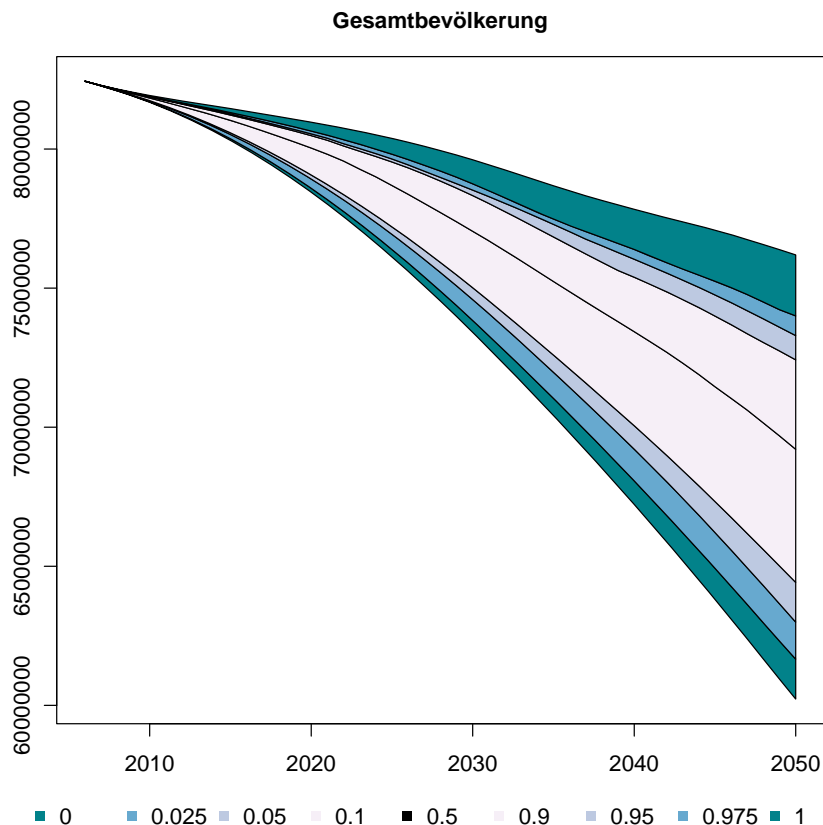


Abbildung 11.2.: Ergebnisverteilung der Gesamtbevölkerung in der Variante II der Beispielprojektion über den Projektionshorizont für die Quantile  $p = 0, 0,025, 0,05, 0,1, 0,5, 0,9, 0,95, 0,975$  und 1

Die erste Graphikart stellt die prognostizierte Entwicklung einer Ergebnisgröße über den

<sup>1</sup>Die Beispielprojektion wird ausführlich im nachfolgenden Kapitel 12 als komplexes Anwendungsbeispiel des PPPM und P3J beschrieben; grundsätzlich wird in der Beispielprojektion die Bevölkerung Deutschlands vom Startjahr 2007 bis zum Zieljahr 2050 mit heterogenen Annahmenverteilungen für nahezu alle (78) Modellparameter prognostiziert, wobei jeweils drei Nachkommengenerationen für die Immigranten und Emigranten berücksichtigt werden.

gesamten Projektionshorizont für ausgewählte (und entsprechend farblich gekennzeichnete) Quantile dar. Folglich können anhand dieser Abbildung die Entwicklung der Prognoseunsicherheit wie auch die Eintrittswahrscheinlichkeiten bestimmter Prognoseergebnisse abgelesen werden.

Die Abbildung 11.2 zeigt so eine Graphik beispielhaft für die Gesamtbevölkerung der Variante II der Beispielprojektion. Daran ist zu erkennen, dass die zukünftige Bevölkerungsentwicklung (unter den getroffenen Annahmen) mit hoher Wahrscheinlichkeit sinken wird und dass die Prognoseunsicherheit mit zunehmender Länge des Projektionshorizonts zunimmt. Im Zieljahr der Prognose wird die Gesamtbevölkerung mit einer Wahrscheinlichkeit von 80 Prozent zwischen 64.421.128 und 72.423.730 Personen liegen.

### 11.2.2. Entwicklung der Wahrscheinlichkeitsverteilung diverser Ergebnisgrößen für ausgewählte Projektionsintervalle

Die zweite Graphikart stellt die Entwicklung der prognostizierten Ergebnisverteilung einer Ergebnisgröße im Querschnitt für einzelne Projektionsintervalle dar. Für jedes ausgewählte Projektionsintervall wird ein Histogramm erstellt, das neben der Ergebnisverteilung auch jeweils den fixen Jump-off Wert der jeweiligen Ergebnisgröße als vertikale rote Linie zeigt. Folglich kann anhand dieser Abbildung die Entwicklung des Unsicherheitsbereichs vom Jump-off Wert der jeweiligen Ergebnisgröße über mehrere Projektionsintervalle analysiert werden.

Die Abbildung 11.3 zeigt dafür beispielhaft die Wahrscheinlichkeitsverteilung der Gesamtbevölkerung von 2010 bis 2050 (im Abstand von jeweils 5 Jahren) für die Variante II der Beispielprojektion. Daran ist zu erkennen, dass die Ergebnisverteilung, ausgehend vom fixen Jump-off Wert aus dem Jahr 2006, mit zunehmender Länge des Projektionshorizonts nach links rutscht und immer breiter wird. Diese Entwicklungen sind gleichbedeutend mit einem Bevölkerungsrückgang und einem Anstieg der Prognoseunsicherheit. Daneben ist zu erkennen, dass die Ergebnisverteilung einer Normalverteilung ähnelt, wobei drei *Entwicklungsszenarien* besonders hervorstechen.

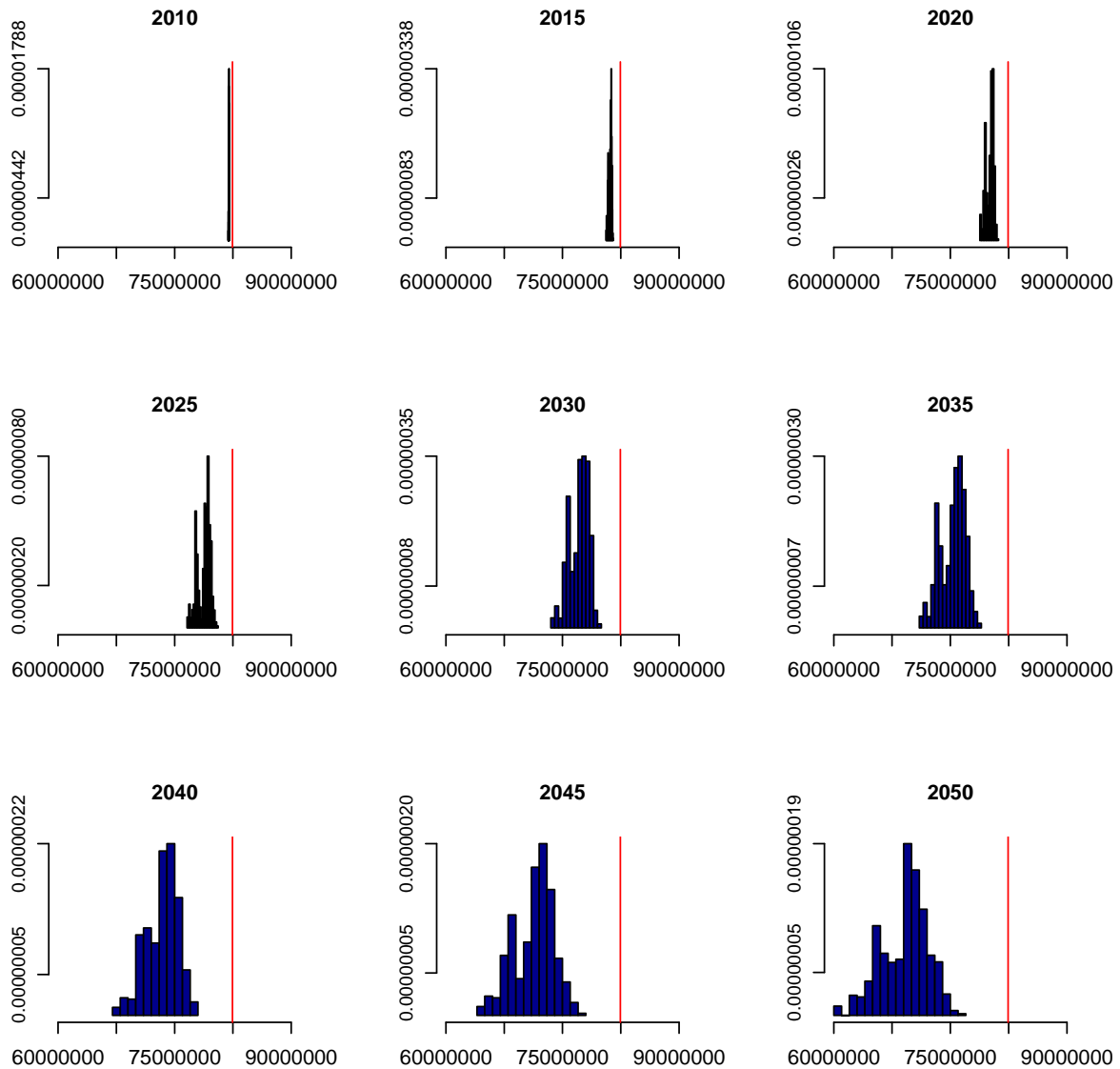


Abbildung 11.3.: Wahrscheinlichkeitsverteilung der Gesamtbevölkerung in der Variante II der Beispielprojektion von 2010 bis 2050 (im Abstand von jeweils 5 Jahren); die rote vertikale Linie repräsentiert den Jump-off Wert der Gesamtbevölkerung im Jahr 2006.

### 11.2.3. Entwicklung der Alters- und Geschlechtsstruktur diverser Ergebnisgrößen

Die dritte Graphikart entspricht einer herkömmlichen Bevölkerungspyramide, die die Alters- und Geschlechtsstruktur einer prognostizierten (Sub)Population in einem bestimmten

Projektionsintervall darstellt. Um die Entwicklung unterschiedlich wahrscheinlicher Prognoseergebnisse darzustellen, werden abermals ausgewählte Quantilswerte für die Bestände der Männer (links) und Frauen (rechts) aufgezeigt; zudem wird noch der Jump-off Bestand als blaue bzw. rote Linie abgebildet, um die Entwicklung der Altersstruktur kenntlich zu machen.

Die Abbildung 11.4 zeigt so eine Bevölkerungspyramide beispielhaft für die Gesamtbevölkerung im Zieljahr 2050 der Variante II der Beispielprojektion. Daran ist zu erkennen, dass zwischen dem Jump-off Jahr 2006 und dem Zieljahr 2050 eine demographische Alterung stattfindet, bei der der Anteil der älteren Personen zunimmt und der Anteil der jungen Personen abnimmt. Darüber hinaus ist der Unsicherheitsbereich besonders in den jungen und hohen Altern ausgeprägt, wohingegen er bei den mittleren Erwachsenenaltern (insbesondere bei den Frauen) relativ eng ist. Folglich gibt es bei dieser Projektion eine größere Unsicherheit bei der Prognose der Sterblichkeit für die höheren Alter wie auch bei der Prognose der Fertilität bzw. der zukünftigen Anzahl der Geburten.

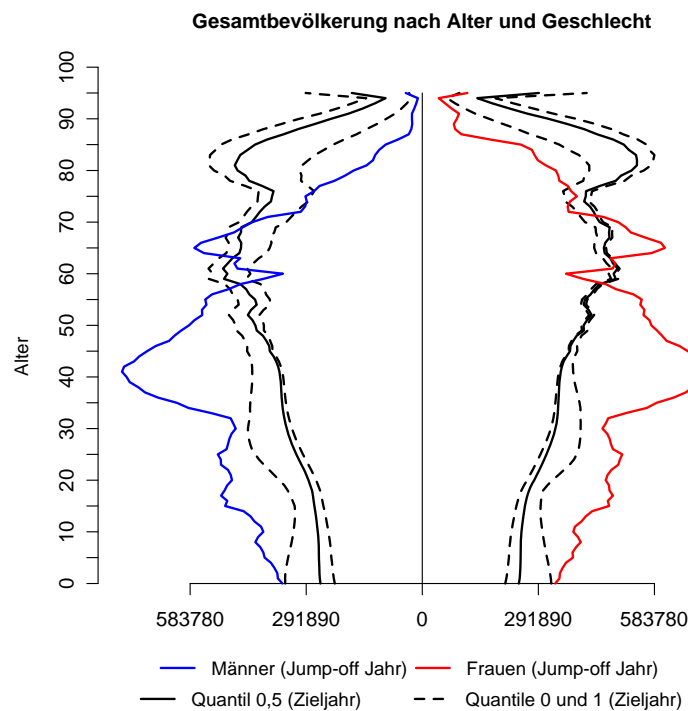


Abbildung 11.4.: Gesamtbevölkerung nach Alter und Geschlecht (Männer links, Frauen rechts) im Jump-off Jahr 2006 (blau und rot) und im Zieljahr 2050 (schwarz) für die Quantile 0 (gestrichelt), 0,5 (durchgezogen) und 1 (gestrichelt) in der Variante II der Beispielprojektion

### 11.3. Erweiterte Ergebnisanalyse

Bei einer erweiterten Ergebnisanalyse kann es erforderlich sein, den Effekt ausgewählter Annahmenverläufe auf die Prognoseergebnisse zu analysieren. Deshalb werden im PPPM für jeden Prognosedurchlauf sowohl die prognostizierten Werte der wesentlichen Ergebnisgrößen als auch die gezogenen Annahmenverläufe aller Modellparameter gespeichert, sodass nicht nur die Ergebnisverteilungen diverser Ergebnisgrößen im Längs- und Querschnitt analysiert werden können. Vielmehr können dadurch auch (im Gegensatz zu herkömmlichen probabilistischen Bevölkerungsprognosemodellen) Ergebnisverläufe extrahiert und zu so genannten *Miniprojektionen* zusammengeschlossen werden, die alle einen oder mehrere bestimmte Annahmenverläufe eines oder mehrerer Modellparameter gemein haben. Außerdem kann mit einem Levelplot graphisch gezeigt werden, welche Annahmenverlaufskombinationen in besonders niedrigen, mittleren oder auch hohen Bevölkerungsbeständen für die (Sub)Populationen resultieren.

#### 11.3.1. Miniprojektionen

Mit *Miniprojektionen* kann demzufolge die originale Ergebnisverteilung (der simulativen und (quasi-)vollständigen Berechnung (vgl. Abschnitt 10.3.2 ab Seite 207)) auf Ergebnisverläufe reduziert werden, die ausschließlich ein oder mehrere Annahmenverläufe eines oder mehrerer Modellparameter gemein haben.

So kann beispielsweise überprüft werden, welchen Effekt Sterblichkeitsannahmen auf die Ergebnisverteilung der (Sub)Populationen haben. In der Variante II der Beispielpjektion gibt es für jede Subpopulation mehrere Annahmenverläufe zur Sterblichkeit, die insgesamt drei erwartbaren Entwicklungsszenarien entsprechen: Die blauen bzw. roten Annahmenverläufe nehmen eine relativ hohe bzw. weniger hohe Sterblichkeitsreduktion an, wohingegen der grüne Annahmenverlauf keinen Sterblichkeitsfortschritt annimmt (vgl. Abschnitt 12.4 ab Seite 242). Daraus können nun drei Miniprojektionen erzeugt werden, die ausschließlich Ergebnisverläufe mit (Theorie- bzw. Erwartungs-)konsistenten Annahmenverläufen zur Sterblichkeit enthalten. So enthalten die blaue, rote bzw. grüne Miniprojektion ausschließlich Annahmenverläufe aus dem blauen, roten bzw. grünen Set zur Sterblichkeit. Somit sind die Annahmen zur Sterblichkeit in jeder Miniprojektion nicht nur (wie im Limited Type) innerhalb eines, sondern zwischen allen Prognosedurchläufen (Theorie- bzw. Erwartungs-) konsistent.

Mit dem Vergleich der Prognoseergebnisse dieser drei Miniprojektionen und der Variante II der Beispielprojektion kann nun ermittelt werden, inwiefern sich die Ergebnisverteilung aufgrund der jeweils eingeschlossenen Annahmen zur Sterblichkeit verändert. Die Abbildungen 11.5 und 11.6 zeigen dazu die Wahrscheinlichkeitsverteilung im Zieljahr wie auch die Ergebnisverteilung über den Projektionshorizont für die Gesamtbevölkerung der Variante II der Beispielprojektion und der drei Miniprojektionen; die Wahrscheinlichkeitsverteilung entspricht dabei jeweils den aufsteigend sortierten Prognoseergebnissen im Zieljahr.

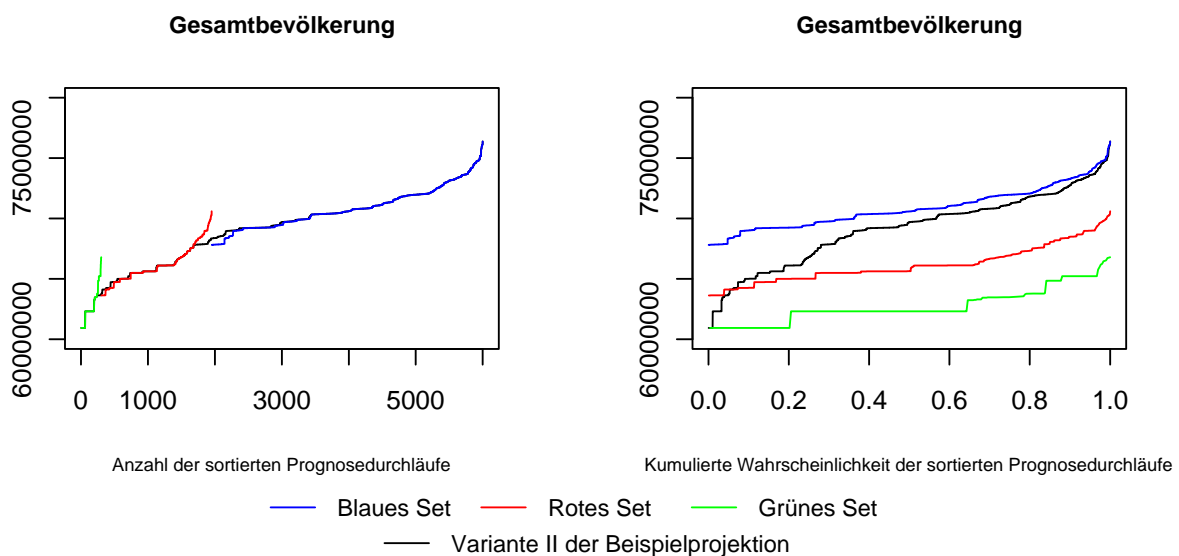


Abbildung 11.5.: Kumulierte Wahrscheinlichkeitsverteilung der Gesamtbevölkerung im Zieljahr der Variante II der Beispielprojektion (schwarz) und ausgewählter Miniprojektionen (blau, rot, grün); die Miniprojektionen enthalten ausschließlich Annahmenverläufe aus dem blauen, roten bzw. grünen Set zur Sterblichkeit von der Variante II der Beispielprojektion.

Es ist in beiden Abbildungen deutlich zu erkennen, dass sich die Wahrscheinlichkeitsverteilung (im Zieljahr) wie auch die Ergebnisverteilung (über den Projektionshorizont) für die Gesamtbevölkerung der Variante II der Beispielprojektion systematisch aus denen der drei Miniprojektionen zusammensetzt. So summieren sich auch die 4.049, 1.649 und 302 Prognosewerte (bzw. Ergebnisverläufe) der blauen, roten und grünen Miniprojektion auf die 6.000 Prognosewerte (bzw. Ergebnisverläufe) der Variante II der Beispielprojektion. Außerdem ist zu erkennen, dass die Prognoseergebnisse der grünen, roten und blauen Miniprojektion jeweils



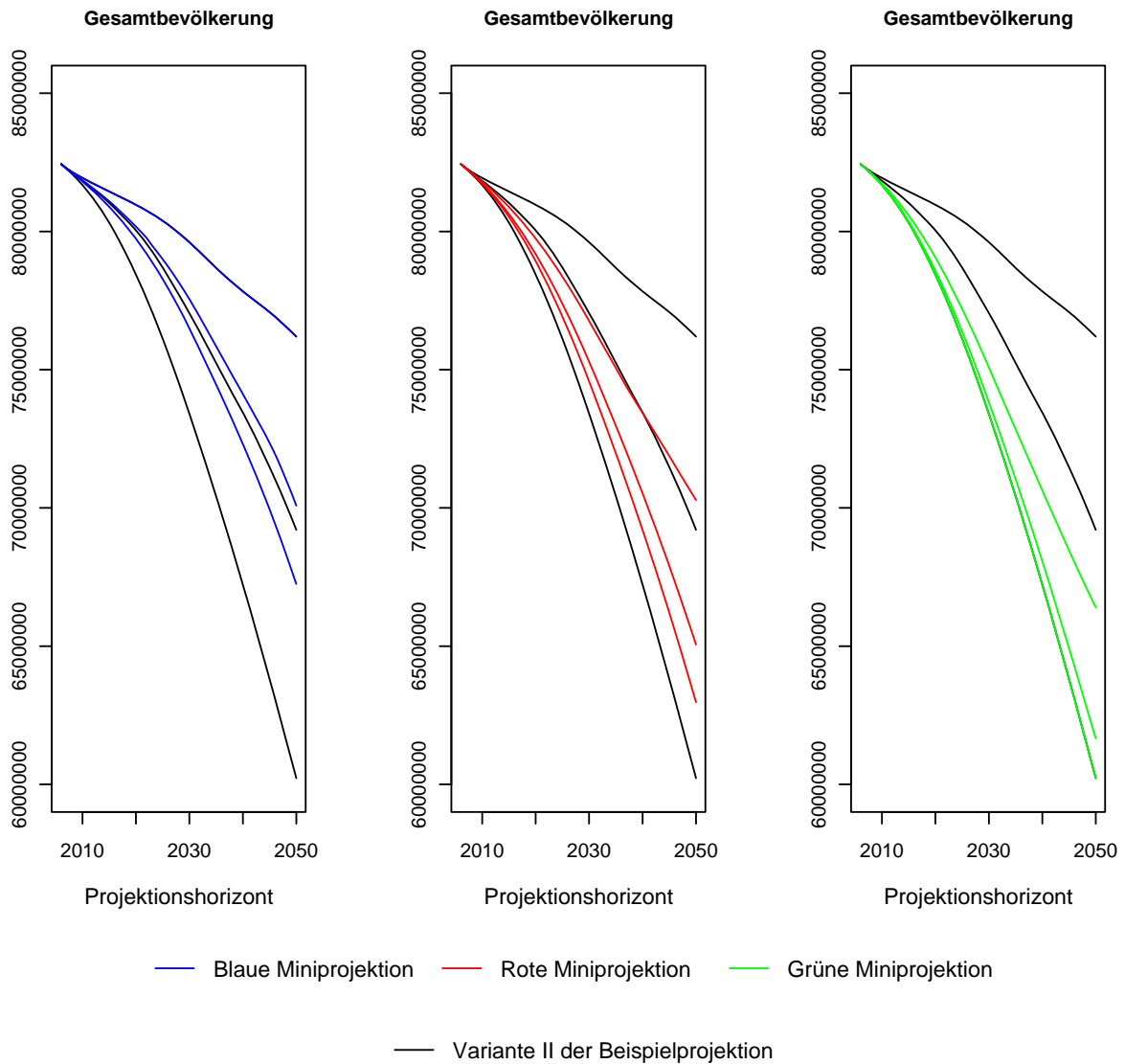


Abbildung 11.6.: Gesamtbevölkerung der Variante II der Beispielprojektion (schwarz) und ausgewählter Miniprojektionen (blau, rot, grün) für die Quantile  $p = 0, 0,5$  und 1 über den Projektionshorizont; die blaue, rote, grüne Miniprojektion enthält ausschließlich Annahmenverläufe mit einer relativ hohen, weniger hohen, ausbleibenden Sterblichkeitsreduktion.

den unteren, mittleren und oberen Bereich der Wahrscheinlichkeits- bzw. Ergebnisverteilung der Variante II der Beispielprojektion abdecken. Folglich hat der Aus- bzw. Einschluss einzelner Annahmenverläufe (zur Sterblichkeit) insbesondere bei einer heterogenen Annahmenverteilung einen wesentlichen Einfluss auf die Ergebnisverteilung. Außerdem kann damit gezeigt

werden, wie die Prognoseunsicherheit durch den Einschluss mehrerer (, heterogener) und subjektiv zu gewichtender Annahmenverläufe im PPPM umfassender gegenüber herkömmlichen Prognosemodellen eingefangen werden kann.

### 11.3.2. Levelplot

Um den Einfluss der Annahmen auf die Höhe der Prognoseergebnisse zu untersuchen, kann im PPPM bzw. P3J aber nicht nur mit den Miniprojektionen die Ergebnisverteilung auf Ergebnisverläufe reduziert werden, die alle ein oder mehrere Annahmenverläufe gemein haben. Obendrein kann z.B. auch noch ein Levelplot erzeugt werden, der die zufällig gezogenen Annahmenverläufe pro Modellparameter für alle aufsteigend sortierten Ergebnisverläufe einer beliebigen Ergebnisgröße darstellt. Damit können dann relativ einfach Annahmenkombinationen erkannt werden, die besonders niedrige, mittlere oder auch hohe Werte der jeweiligen Ergebnisgröße erzeugen.

Um so einen Levelplot beispielhaft für die prognostizierte Gesamtbevölkerung der blauen Miniprojektion (vgl. Abschnitt 11.3.1) zu erstellen, werden zunächst alle (zufällig gezogenen) Annahmenverläufe der insgesamt 78 Modellparameter (vgl. Tabelle 11.2) farblich kodiert; in diesem Beispiel werden die bereits verwendeten Farbnamen (grün, rot, dunkelrot, blau, türkis und dunkelblau) der Annahmenverläufe auch zur Bestimmung ihrer (Linien)Farben im Levelplot genutzt; sofern es allerdings nur einen Annahmenverlauf pro Modellparameter gibt, erhält dieser die Farbe schwarz.

Die Abbildung 11.7 zeigt nun den Levelplot für die prognostizierte Gesamtbevölkerung der blauen Miniprojektion. Dabei ist vordergründig zu erkennen, dass ausschließlich Annahmenverläufe aus dem blauen Set zur Sterblichkeit gezogen worden sind. Bei genauerer Betrachtung fällt zudem auf, dass sich die einschränkende Vorgabe der Variante II der Beispielprojektion, jeweils nur ähnliche Annahmenverläufe zur Mortalität, Fertilität und Migration zu kombinieren, im Levelplot dadurch ausdrückt, dass sich Farbabschnitte (grün, rot oder blau) für in einem Settyp zusammengeschlossene Modellparameter in jedem Prognosedurchlauf gleichen; so wird beispielsweise ein Annahmenverlauf aus dem roten Set für alle neun im Settyp Fertilität eingeschlossenen Modellparameter (mit den Nummern 8, 18, 28, 38, 46, 54, 62, 70 und 78) in den sortierten Ergebnisverläufen 193 bis 664 gezogen. Außerdem ist zu erkennen, dass niedrige, mittlere und hohe Prognosewerte für die Gesamtbevölkerung mit grünen, roten und vermehrt blauen Annahmenverläufen zur Fertilität und Migration (Modellparameter

1	$p_{95+}^{E,m}$	27	$l_x^{SZ,w}$	53	$l_x^{SFzG(2),w}$
2	$p_{95+}^{E,w}$	28	$f_x^{SZ}$	54	$f_x^{SFzG(2)}$
3	$\pi(0,5D_0)^{E,m}$	29	$Z^m$	55	$p_{95+}^{SZG(2),m}$
4	$\pi(0,5D_0)^{E,w}$	30	$Z^w$	56	$p_{95+}^{SZG(2),w}$
5	$\pi(P_0)^{E,m}$	31	$p_{95+}^{SFzG(1),m}$	57	$\pi(0,5D_0)^{SZG(2),m}$
6	$l_x^{E,m}$	32	$p_{95+}^{SFzG(1),w}$	58	$\pi(0,5D_0)^{SZG(2),w}$
7	$l_x^{E,w}$	33	$\pi(0,5D_0)^{SFzG(1),m}$	59	$\pi(P_0)^{SZG(2),m}$
8	$f_x^E$	34	$\pi(0,5D_0)^{SFzG(1),w}$	60	$l_x^{SZG(2),m}$
9	$P_{31.12.Jump-offJahr}^{E,m}$	35	$\pi(P_0)^{SFzG(1),m}$	61	$l_x^{SZG(2),w}$
10	$P_{31.12.Jump-offJahr}^{E,m}$	36	$l_x^{SFzG(1),m}$	62	$f_x^{SZG(2)}$
11	$p_{95+}^{SFz,m}$	37	$l_x^{SFzG(1),w}$	63	$p_{95+}^{SFzG(3),m}$
12	$p_{95+}^{SFz,w}$	38	$f_x^{SFzG(1)}$	64	$p_{95+}^{SFzG(3),w}$
13	$\pi(0,5D_0)^{SFz,m}$	39	$p_{95+}^{SZG(1),m}$	65	$\pi(0,5D_0)^{SFzG(3),m}$
14	$\pi(0,5D_0)^{SFz,w}$	40	$p_{95+}^{SZG(1),w}$	66	$\pi(0,5D_0)^{SFzG(3),w}$
15	$\pi(P_0)^{SFz,m}$	41	$\pi(0,5D_0)^{SZG(1),m}$	67	$\pi(P_0)^{SFzG(3),m}$
16	$l_x^{SFz,m}$	42	$\pi(0,5D_0)^{SZG(1),w}$	68	$l_x^{SFzG(3),m}$
17	$l_x^{SFz,w}$	43	$\pi(P_0)^{SZG(1),m}$	69	$l_x^{SFzG(3),w}$
18	$f_x^{SFz}$	44	$l_x^{SZG(1),m}$	70	$f_x^{SFzG(3)}$
19	$Fz^m$	45	$l_x^{SZG(1),w}$	71	$p_{95+}^{SZG(3),m}$
20	$Fz^w$	46	$f_x^{SZG(1)}$	72	$p_{95+}^{SZG(3),w}$
21	$p_{95+}^{SZ,m}$	47	$p_{95+}^{SFzG(2),m}$	73	$\pi(0,5D_0)^{SZG(3),m}$
22	$p_{95+}^{SZ,w}$	48	$p_{95+}^{SFzG(2),w}$	74	$\pi(0,5D_0)^{SZG(3),w}$
23	$\pi(0,5D_0)^{SZ,m}$	49	$\pi(0,5D_0)^{SFzG(2),m}$	75	$\pi(P_0)^{SZG(3),m}$
24	$\pi(0,5D_0)^{SZ,w}$	50	$\pi(0,5D_0)^{SFzG(2),w}$	76	$l_x^{SZG(3),m}$
25	$\pi(P_0^{SZ,m})$	51	$\pi(P_0^{SFzG(2),m})$	77	$l_x^{SZG(3),w}$
26	$l_x^{SZ,m}$	52	$l_x^{SFzG(2),m}$	78	$f_x^{SZG(3)}$

Tabelle 11.2.: Reihenfolge der 78 Modellparameter im Levelplot (der Abbildung 11.7) für die Gesamtbevölkerung der blauen Miniprojektion;  $E$ ,  $w$  und  $m$  stehen für *Einheimische*, *weiblich* und *männlich*.

mit den Nummern 19, 20, 29 und 30) korrelieren; bei der Sterblichkeit sind hingegen die Beziehungen zwischen den zufällig gezogenen Annahmenverläufen und der Höhe der prognostizierten Gesamtbevölkerung weniger stark ausgeprägt; hier ist nur bei der Sterblichkeit der

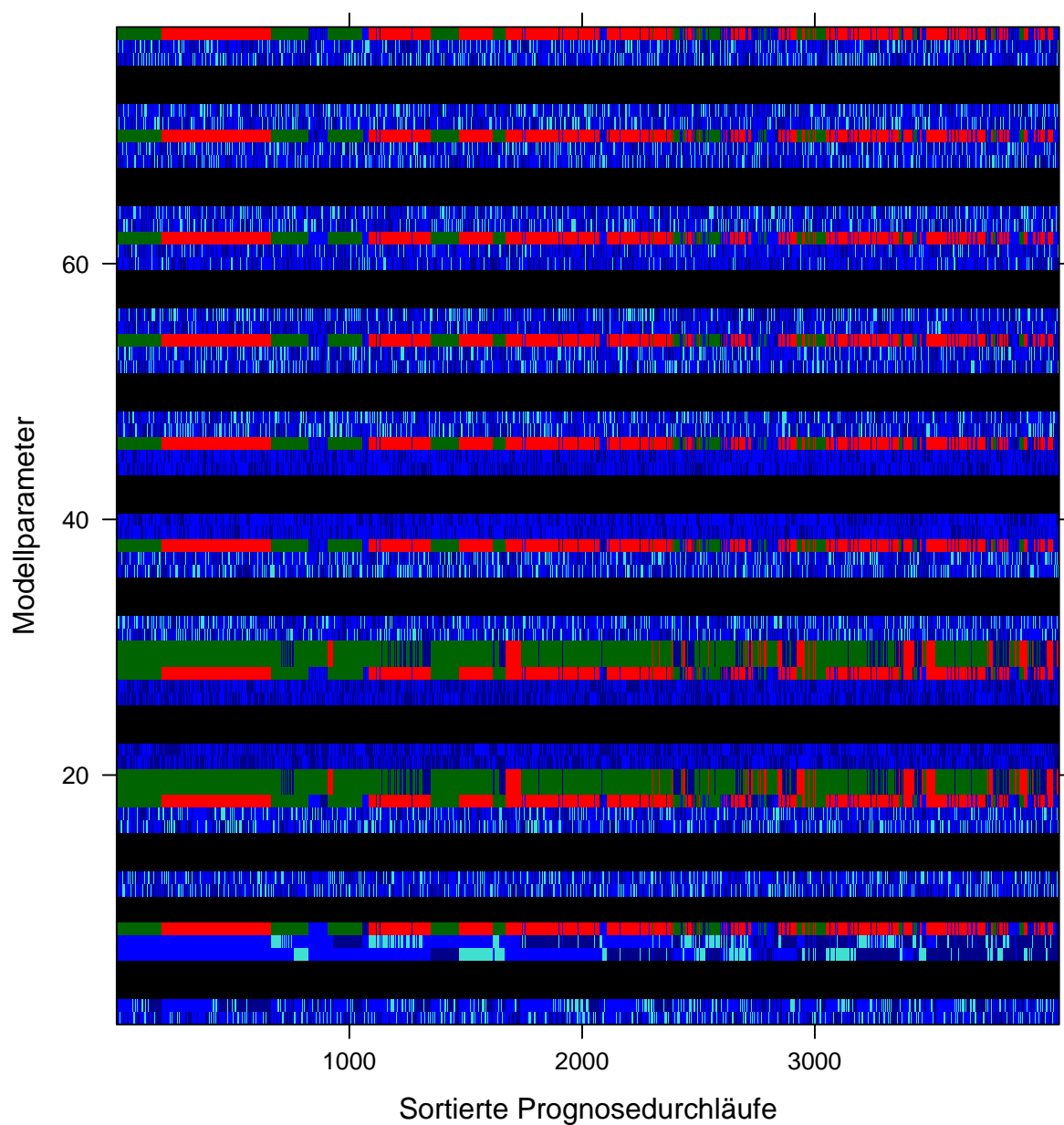


Abbildung 11.7.: Levelplot mit den zufällig gezogenen, farblich codierten (dunkelblau, blau, türkis, rot, grün) Annahmenverläufen aller 78 Modellparameter für alle aufsteigend sortierten Prognosewerte der Gesamtbevölkerung in der blauen Miniprojektion aus dem Abschnitt 11.3.1

Einheimischen (Modellparameter mit den Nummern 6 und 7) zu erkennen, dass eine niedrige, mittlere und höhere Gesamtbevölkerung mit den blauen und dann vermehrt mit den türkisen und dunkelblauen Annahmenverläufen prognostiziert wird, wobei der dunkelblaue

Annahmenverlauf auch die größte Sterblichkeitsreduktion annimmt.

## 11.4. P3J — Ausgabe und Zusammenstellung der Prognoseergebnisse

Die Ergebnisse einer Prognose werden, analog zu den Eingabedaten (vgl. Abschnitt 9.3.2 ab Seite 179), in einer Datenbank gespeichert. Die Speicherung der Ergebniswerte mit den sie erzeugenden Annahmenwerten in jedem Prognosedurchlauf erweitert die Ergebnisanalyse, sodass nicht nur die Ergebniswerte allein, sondern eben zusammen mit den sie erzeugenden Annahmen z. B. in Miniprojektionen (vgl. Abschnitt 11.3.1) und Levelplots (vgl. Abschnitt 11.3.2) analysiert werden können.

Grundsätzlich bietet das P3J zwei Möglichkeiten an, um die generierten Prognoseergebnisse zu verwenden: Einerseits können die Ergebniswerte als *Rohdaten* gespeichert oder automatisch aggregiert (vgl. Tabelle 11.1) und zu einem *Ergebnisreport* zusammengestellt werden. Die folgenden Abschnitte erläutern die Funktionsweise dieser beiden Alternativen.

### 11.4.1. Datenexport

Ein Export der Ergebniswerte in unbearbeiteter Form erlaubt ihre flexible Weiterverarbeitung mit Hilfe anderer Software. Das P3J exportiert Daten in Form von komma-separierten Werten, kurz CSV (*comma-separated values*). Dieses Dateiformat wird von vielen Programmen zur Datenanalyse unterstützt. Durch den Export der Rohdaten ist jeder Nutzer in der Lage, die Ergebniswerte (der wesentlichen Ergebnisgrößen) aus jedem Prognosedurchlauf mit selbst gewählten Methoden zu untersuchen.

Für den Export wählt der Nutzer ein Zielverzeichnis, in das die Rohdaten gespeichert werden sollen. In diesem Verzeichnis wird nun für jeden Prognosedurchlauf ein neues Unterverzeichnis namens *trial\_x* angelegt, wobei  $x = 1, \dots, n$  die Nummer des jeweiligen Prognosedurchlaufs (in der Reihenfolge der Ausführung) angibt. Die komma-separierten Werte werden in Dateien mit entsprechenden Namen abgelegt, z. B. sind in der Datei *natives\_end\_x.m.csv* im Unterverzeichnis *trial\_4* die im 4. Prognosedurchlauf berechneten Jahresendbestände der einheimischen Männer nach Alter gespeichert.

### 11.4.2. Reportgenerierung

Die Generierung eines Ergebnisreports umfasst mehrere Schritte. Zunächst wird, wie beim Datenexport, ein Zielverzeichnis für alle Dateien des Reports angegeben. In diesem Zielverzeichnis befinden sich dann die Dateien, die zur Erzeugung des Reports notwendig sind; dazu zählt z.B. auch das Unterverzeichnis *aggregated*, das die aggregierten Ergebniswerte zur Erstellung des Reports enthält. Über einen Filter kann der Nutzer zusätzlich bestimmen, ob die Ergebniswerte aller oder nur ausgewählter Ergebnisverläufe exportiert bzw. aggregiert werden sollen. Dies geschieht, indem der Nutzer eine Zeichenkette angibt und dann auswählt, ob die Zeichenkette im Namen eines (oder mehrerer) ausgewählter Sets oder Annahmenverläufe vorkommen muss (bzw. für  $\dot{\sigma} \in \dot{S}$  oder  $\dot{a} \in \dot{A}_{\dot{S}}$  gilt; vgl. dazu Abschnitt 10.3.1). Sofern eindeutige Namen bei der Dateneingabe (vgl. Abschnitt 9.3.2) verwendet worden sind, lassen sich mit dieser Funktion die in Abschnitt 11.3.1 beschriebenen Miniprojektionen erstellen bzw. extrahieren. Wegen des zusätzlichen Filtermechanismus und des hohen Speicherverbrauchs für die verschiedenen Ergebnisgrößen (vgl. Tabelle 11.1) findet die Aggregation der Daten erst statt, *nachdem* der Nutzer die Erzeugung des Reports gestartet hat.

Die aggregierten Daten werden, analog zu den Rohdaten, in CSV-Dateien geschrieben. Zusätzlich wird für jeden Trial die Annahmenverlaufskombination  $\dot{A}_{\dot{S}}$  sowie deren Eintrittswahrscheinlichkeit  $Pr[\dot{A}_{\dot{S}}]$  (vgl. Gleichung 10.38 auf Seite 206) in einer nach Eintrittswahrscheinlichkeit sortierten Liste gespeichert. Diese Daten können, wie in Abschnitt 11.3.2 beschrieben, mit einem Levelplot visualisiert werden.

Der Report selbst wird nur zum Teil durch das P3J erzeugt, sodass unterstützend das Textsatzsystem  $\text{\LaTeX}^2$  und die Statistik-Umgebung  $R^3$  verwendet werden.  $\text{\LaTeX}$  ist ein weit verbreitetes Werkzeug zur Erstellung von (meist wissenschaftlichen) Texten. Es kann Dokumente in vielen verschiedenen Formaten, z.B. PDF oder HTML, erzeugen.  $R$  ist eine Programmiersprache und -umgebung für statistische Berechnungen, und ebenfalls weit verbreitet. Sowohl  $\text{\LaTeX}$  als auch  $R$  sind kostenlos erhältlich und damit jedem Nutzer des P3J leicht zugänglich. Die Kombination von  $\text{\LaTeX}$  und  $R$  wird durch das ebenfalls frei erhältliche  $R$ -Paket SWEAVE [161] ermöglicht, mit dem  $R$ -Befehle innerhalb eines  $\text{\LaTeX}$ -Dokuments ausgeführt und deren Ergebnisse direkt in das Dokument eingebettet

---

<sup>2</sup><http://www.latex-project.org>, letzter Zugriff 12/2010

<sup>3</sup><http://www.r-project.org>, letzter Zugriff 12/2010

werden können.

Sobald mit dem P3J eine vorläufige Version des Ergebnisreports generiert worden ist, müssen zwei zusätzliche Schritte erfolgen, um das endgültige Resultat — den vollständigen Ergebnisreport (z. B. als PDF-Datei) — zu erhalten: So muss zunächst in R SWEAVE [161] auf der vom P3J erzeugten Reportdatei (`report.Rtex`) ausgeführt werden, um anschließend mit  $\text{\LaTeX}$  aus der daraus resultierenden Reportdatei (`report.tex`) ein PDF-Dokument mit dem vollständigen Ergebnisreport (`report.pdf`) zu erzeugen.

Die Reportdatei `report.Rtex` wird mit einer weiteren frei erhältlichen Zusatzbibliothek, FREEMARKER,<sup>4</sup> erzeugt. FREEMARKER stellt Funktionen bereit, um die Vorlage eines Ergebnisreports mit Verweisen auf die im Verzeichnis *aggregated* liegenden CSV-Dateien zu füllen. Gleichzeitig fügt FREEMARKER alle für die Erzeugung des Reports benötigten R-Befehle in die Vorlage. Somit ist es z. B. möglich, ohne Aufwand für den Nutzer viele verschiedene Plots für alle in der Prognose betrachteten Subpopulationen zu erzeugen.

Die grundlegende Vorgehensweise zur Generierung eines Ergebnisreports zeigt die Abbildung 11.8. Wie bereits in Abschnitt 9.3.2 (ab Seite 179) beschrieben, wird die Persistenzschnittstelle HIBERNATE zur Verwaltung der in der Datenbank gespeicherten Annahmenverläufe und Ergebnisgrößen verwendet. Letztere werden für die Reportgenerierung ausgelesen, aggregiert und in CSV-Dateien geschrieben. Danach erzeugt FREEMARKER eine Reportdatei (`report.Rtex`) mit Verweisen auf die angelegten CSV-Dateien und R-Befehle, die diese darstellen bzw. analysieren sollen. Diese Schritte werden automatisch vom P3J ausgeführt; der Nutzer muss danach jedoch zunächst R und dann  $\text{\LaTeX}$  (siehe weiße Kästen in der Abbildung 11.8) auf die generierte Datei anwenden, um den vollständigen Report (`report.pdf`) zu erhalten. Diese Vorgehensweise mag recht umständlich erscheinen, hat aber einige Vorteile. So ist es leicht möglich, Farben und andere Eigenschaften eines bestimmten Plots zu ändern, indem der Nutzer den entsprechenden R-Code in der vom P3J generierten Datei (`plotting.R`) ändert. Ebenso können in den Report wichtige Zusatzinformationen, z. B. bezüglich grundlegender Annahmen oder verwendeter Daten, zusammen mit den Ergebnissen der Prognose dargestellt werden, indem sie zusätzlich in das  $\text{\LaTeX}$ -Dokument eingebunden werden.

---

<sup>4</sup><http://freemarker.sourceforge.net>, letzter Zugriff 12/2010

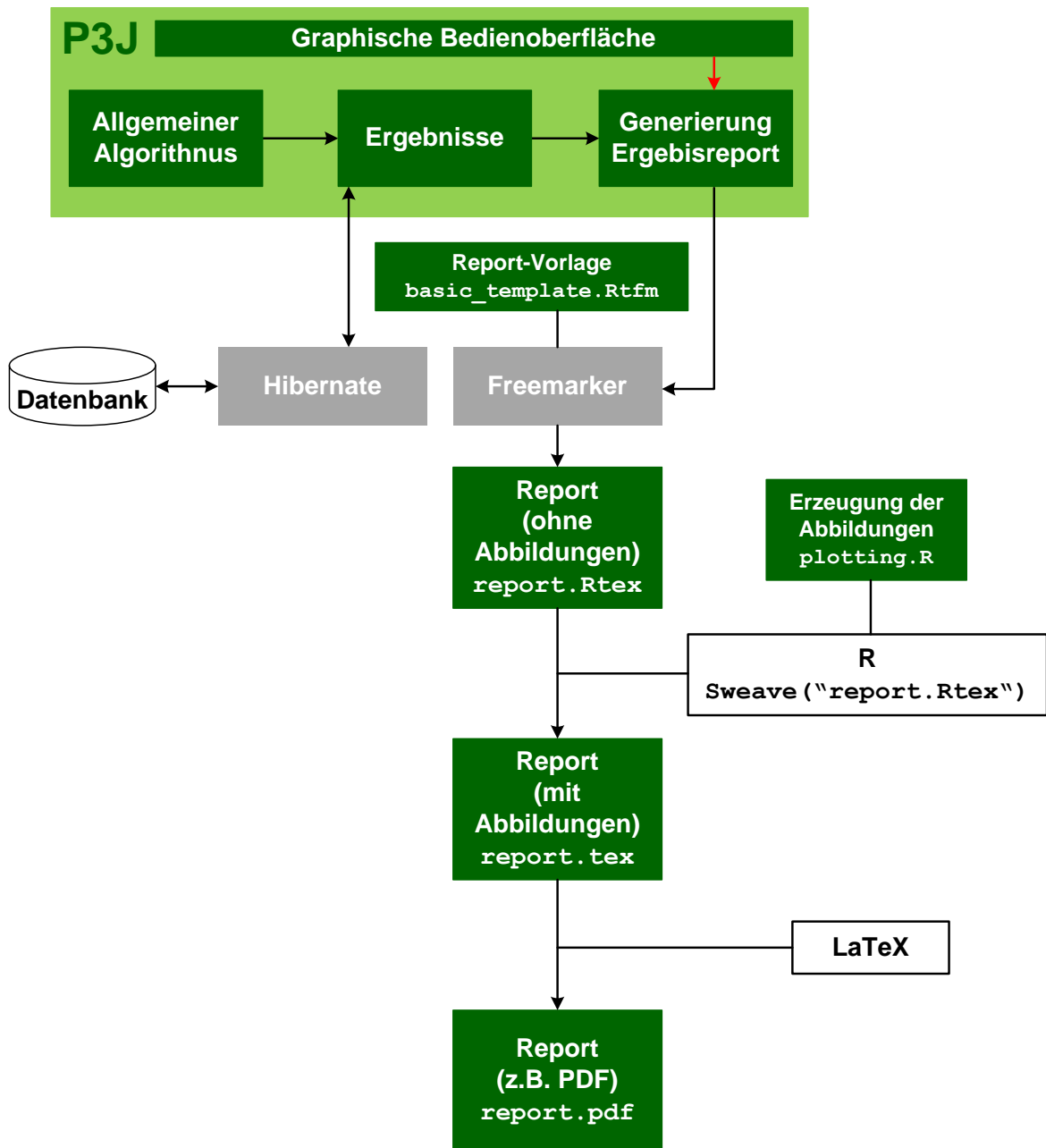


Abbildung 11.8.: Modul für die Generierung eines Ergebnisreports: Freemarker wird verwendet, um aus einer Vorlagendatei eine Reportdatei zu generieren; diese muss dann vom Nutzer noch mit R und  $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$  zu einem vollständigen Dokument verarbeitet werden.



## 12. Beispielhafte Anwendung des PPPM und P3J

Nachdem das PPPM und seine prototypische Realisierung P3J in den drei vorherigen Kapiteln formal erläutert worden sind, sollen sie nun in diesem Kapitel beispielhaft angewendet werden.

Um das Prozedere einer probabilistischen Bevölkerungsprognose mit dem PPPM und P3J aufzuzeigen, wird deshalb nachfolgend eine Beispielprojektion mit realen Daten zur Bevölkerung Deutschlands durchgeführt. Die Struktur folgt dabei den einzelnen Arbeitsschritten eines Bevölkerungsprognoseprozesses aus dem Abschnitt 2.2 (ab Seite 18).

### 12.1. Fragestellung bzw. Prognosegegenstand

In der Beispielprojektion soll die Bevölkerung Deutschlands nach Einzelalter (bis zu der nach oben offenen Altersgruppe 95+) und Geschlecht über den Projektionshorizont von 2007 bis 2050 prognostiziert werden. Da sich der Projektionshorizont über 44 Jahre erstreckt, sollen jeweils drei Nachkommengenerationen der direkten Zu- und Fortzüge berücksichtigt werden.

Das vornehmliche Ziel dieser Beispielprojektion besteht darin, die Arbeitsweise und Vorzüge des PPPM und P3J hinsichtlich

- der detaillierten Modellparameter bzw. des Aufbaus nach Subpopulationen
- der Generierung der Annahmenverteilung für jeden Modellparameter
- der Kombination der Annahmenverläufe
- der alternativen Berechnungsmodi und
- der Ergebnisanalyse

aufzuzeigen. Dazu wird die ursprüngliche Beispielprojektion entsprechend variiert und gegebenenfalls mit herkömmlichen Bevölkerungsprognose(modelle)n verglichen.

## 12.2. Auswahl eines geeigneten Prognosemodells

Zur Berechnung der Beispielprojektion (und ihrer Variationen) wird das P3J, die prototypische Realisierung des PPPM als Software, genutzt.

## 12.3. Sammlung relevanter Informationen/Daten

In den Kapiteln 5 bis 7 sind verschiedene Methoden zur Generierung der Annahmen zur Mortalität, Fertilität und Migration beschrieben worden. Ein wesentliches Unterscheidungsmerkmal dieser Methoden ist die Vielfalt der zu verwendenden Informationsquellen. Wissen bzw. Erwartungen über potentielle zukünftige Entwicklungen eines Modellparameters können beispielsweise aus der vergangenen empirischen Entwicklung und/oder entsprechenden Theorien, wissenschaftlichen Diskursen und Erhebungen resultieren. Grundsätzlich ist es sinnvoll, Annahmen vornehmlich auf die gesicherte vergangene Entwicklung zu stützen, allerdings sollten mit zunehmender Länge des Projektionshorizonts zunehmend zusätzliche Informationen korrigierend eingesetzt werden, um auch (zum Zeitpunkt der Prognoseerstellung) unerwartete bzw. unwahrscheinliche Trendumkehrungen, die nicht im Langzeittrend enthalten sind, einfangen zu können.

Demzufolge wird die vergangene empirische Entwicklung der Modellparameter die ausschlaggebende Informationsquelle bei der Annahmengenerierung sein, wobei aber auch zusätzliche Erkenntnisse aus entsprechenden Theorien, wissenschaftlichen Diskursen und anderen Bevölkerungsprognosen für Deutschland (korrigierend) eingesetzt werden. So wird u. a. auf Annahmen zur Mortalität, Fertilität und Migration aus der *12. koordinierten Bevölkerungsprognose* vom Statistischen Bundesamt [266] zurückgegriffen.

## 12.4. Generierung der Annahmen für die Modellparameter

Um nun für jede einzeln zu prognostizierende Subpopulation (vgl. Abschnitt 9.1) eine heterogene Annahmenverteilung aus verschiedenen (und subjektiv zu gewichtenden) Annahmenverläufen (vgl. Abschnitt 10.2.1) zur Mortalität, Fertilität und — wenn erforderlich —

auch zur Migration (extern) zu erstellen, wird ein Methodenmix verwendet, bei dem die Fortschreibung des in der Vergangenheit beobachteten Trends mehr oder weniger stark abgeändert wird, um die jeweils resultierenden Annahmenwerte an eine konkrete Erwartungshaltung bzw. Theorie anzupassen.<sup>1</sup>

#### 12.4.1. Mortalität

##### Methode

Die Annahmenverläufe für den Modellparameter *Überlebende im Alter  $x$*  werden für jede Subpopulation über die Prognose der (logarithmierten) altersspezifischen Sterbewahrscheinlichkeiten mit einem Methodenmix generiert. Diese altersspezifischen Sterbewahrscheinlichkeiten werden mit der Modellgleichung des Ansatzes von Lee und Carter (vgl. Abschnitt 5.2.1 ab Seite 84) (logarithmiert) prognostiziert, wobei die  $\beta_x$  zunächst mit der Singulärwertzerlegung aus den Basisdaten von 1992 bis 2006 geschätzt und anschließend mit Expertenmeinung korrigiert werden.<sup>2</sup> Die  $k_t$  werden (ausgehend von den mit der Singulärwertzerlegung geschätzten Werten für die Basisperiode) mit einem autoregressiven Modell (der Ordnung 1) und einem subjektiv festzulegenden Drift fortgeschrieben.

##### Bestimmung der Jump-off Werte

Um nun mit diesem Methodenmix Sterblichkeitsannahmen zu generieren, müssen zunächst geeignete Jump-off Werte zur altersspezifischen Sterblichkeit für jede Subpopulation ermittelt werden.

Grundsätzlich liegen altersspezifische Sterbewahrscheinlichkeiten für die einheimische und ausländische Bevölkerung aus der amtlichen Statistik für das Berichtsjahr 2006 vor. Im Abschnitt 7.1.2 (ab Seite 141) ist aber bereits erläutert worden, dass die Sterblichkeit der ausländischen Bevölkerung nicht mit der Sterblichkeit der direkten Zuzüge und deren erster (und weiterer) Nachkommengeneration(en) übereinstimmt, sodass diese entsprechend ange-

---

<sup>1</sup>An dieser Stelle sei nochmals betont, dass die Beispielprojektion nicht die wahrscheinlichste zukünftige Bevölkerungsentwicklung prognostiziert, sondern die Arbeitsweise und Vorzüge des PPPM und P3J (gegenüber herkömmlichen Bevölkerungsprognose(modelle)n) zeigen soll. Folglich entsprechen die Annahmenverteilungen nicht einem wirklichkeitsgetreuen Abbild aller potentiellen zukünftigen Entwicklungen, sondern einem Ausschnitt davon.

<sup>2</sup>Die Basisperiode erstreckt sich nur über einen relativ kurzen Zeitraum, um die aktuellen Trends in der Sterblichkeit einzufangen und fortzuschreiben.

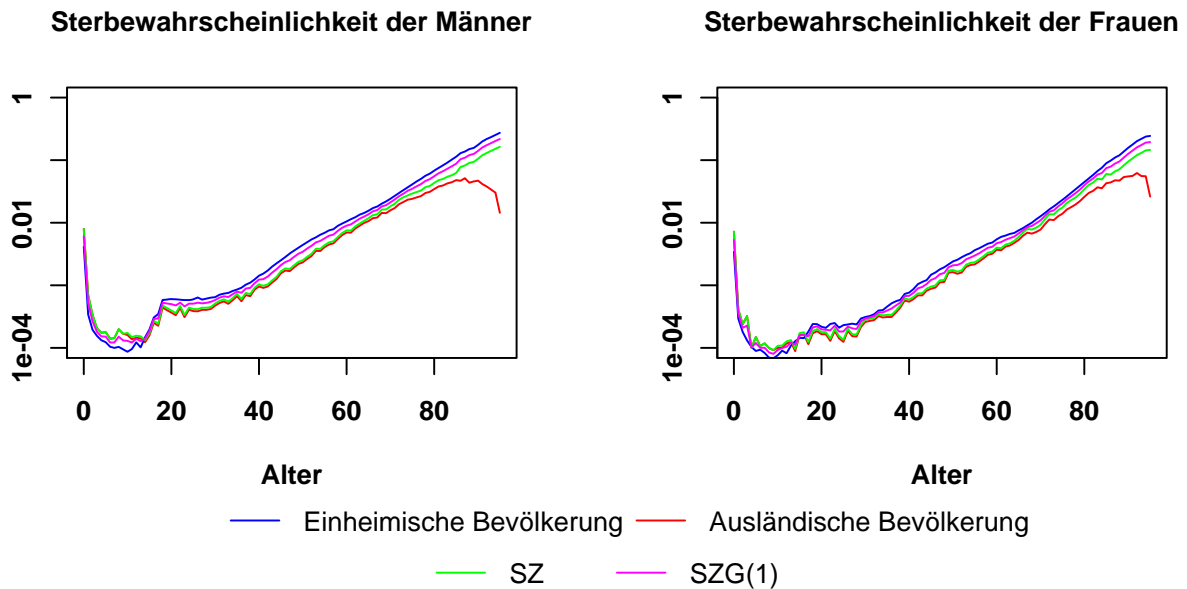


Abbildung 12.1.: Bestimmung der Jump-off Werte für die altersspezifischen Sterbewahrscheinlichkeiten der direkten Zuzüge (SZ) und ihrer ersten Nachkommengeneration (SZG(1)) aus den altersspezifischen Sterbewahrscheinlichkeiten der einheimischen und ausländischen Bevölkerung des Jahres 2006; Datenquelle: Statistisches Bundesamt

passt werden muss. Um nun die Unterschiede in der gemessenen Sterblichkeit zwischen der amtlichen Statistik und dem Ausländerzentralregister wie auch zwischen den Einheimischen, der ausländischen Bevölkerung und den direkten Zuzügen angemessen einzufangen, werden für die direkten Zuzüge (in Anlehnung an Kohls [143]) die Sterbewahrscheinlichkeiten der ausländischen Bevölkerung (aus der amtlichen Statistik) im Alter 0 bis 9 direkt übernommen und im Alter 10 bis 85 mit einem Faktor multipliziert, der bis zum Alter 64 konstant bei 1,08 liegt und sich dann linear bis auf 1,4 für die Frauen bzw. 1,3 für die Männer erhöht. Ab dem Alter 85 wird die Sterblichkeit der direkten Zuzüge aus dem Produkt der Sterblichkeit der Einheimischen und dem Faktor 0,6 ermittelt. Damit haben die direkten Zuzüge (gegenüber den Einheimischen) eine leicht erhöhte Säuglings- und Kindersterblichkeit und eine (leicht) niedrigere Sterblichkeit im Erwachsenenalter, wobei der enorme Sterblichkeitsvorteil insbesondere in den sehr hohen Altern (, wie er laut den Daten der amtlichen Statistik ausgewiesen wird,) abgeschwächt wird. Die Sterblichkeit der ersten Nachkommengeneration der direkten Zuzüge wird dann als arithmetisches Mittel aus der Sterblichkeit der direkten Zuzüge und der einheimischen Bevölkerung ermittelt (vgl. Abbildung 12.1).

Ab der zweiten Nachkommengeneration werden kaum noch Unterschiede zur Sterblichkeit der Einheimischen erwartet, sodass hier dieselben altersspezifischen Sterbewahrscheinlichkeiten als Jump-off Werte herangezogen werden.

### Vergangene Entwicklung und Variationen des Langzeittrends

**Einheimische** Die Abbildung 12.2 zeigt sowohl die vergangene Entwicklung als auch jeweils 6 verschiedene Annahmenverläufe zur zukünftigen Entwicklung der Sterblichkeit mit den subjektiv zugeordneten Eintrittswahrscheinlichkeiten für die einheimischen Männer und Frauen anhand des Parameters *Lebenserwartung bei Geburt*. Zudem enthält die Tabelle 12.1 die konkreten Annahmenwerte zur Lebenserwartung bei Geburt im Zieljahr wie auch die entsprechenden Eintrittswahrscheinlichkeiten.

Annahmenverlauf $j$	Frauen		Männer	
	$e_0$	$pr_j$	$e_0$	$pr_j$
Dunkelblau	90,41	0,2	88,26	0,2
Türkis	89,68	0,1	87,17	0,1
Blau	88,41	0,3	85,32	0,3
Dunkelrot	85,79	0,2	81,68	0,2
Rot	84,76	0,15	80,32	0,15
Grün	82,03	0,05	76,86	0,05

Tabelle 12.1.: Annahmenverläufe  $j$  zur Mortalität der einheimischen Frauen und Männer charakterisiert durch die *Lebenserwartung bei Geburt*  $e_0$  im Zieljahr und die subjektiv zugeordnete Eintrittswahrscheinlichkeit  $pr_j$ .

Darüber hinaus zeigt die Abbildung 12.3 ergänzend den dunkelblauen, roten und grünen Annahmenverlauf der (logarithmierten) altersspezifischen Sterbewahrscheinlichkeiten und der daraus eindeutig zu berechnenden *Überlebenden im Alter  $x$*  für die einheimischen Männer nach Einzelalter über die Basisperiode (von 1992 bis 2006) und den Projektionshorizont (von 2007 bis 2050).

In der Vergangenheit ist die Sterblichkeit kontinuierlich gesunken, sodass die Lebenserwartung bei Geburt linear um etwa 3 Monate pro Jahr bzw. 2,5 Jahre pro Dekade gestiegen ist (vgl. Abschnitt 5.2.3 ab Seite 98). Die Fortführung dieses Langzeittrends in die Zukunft

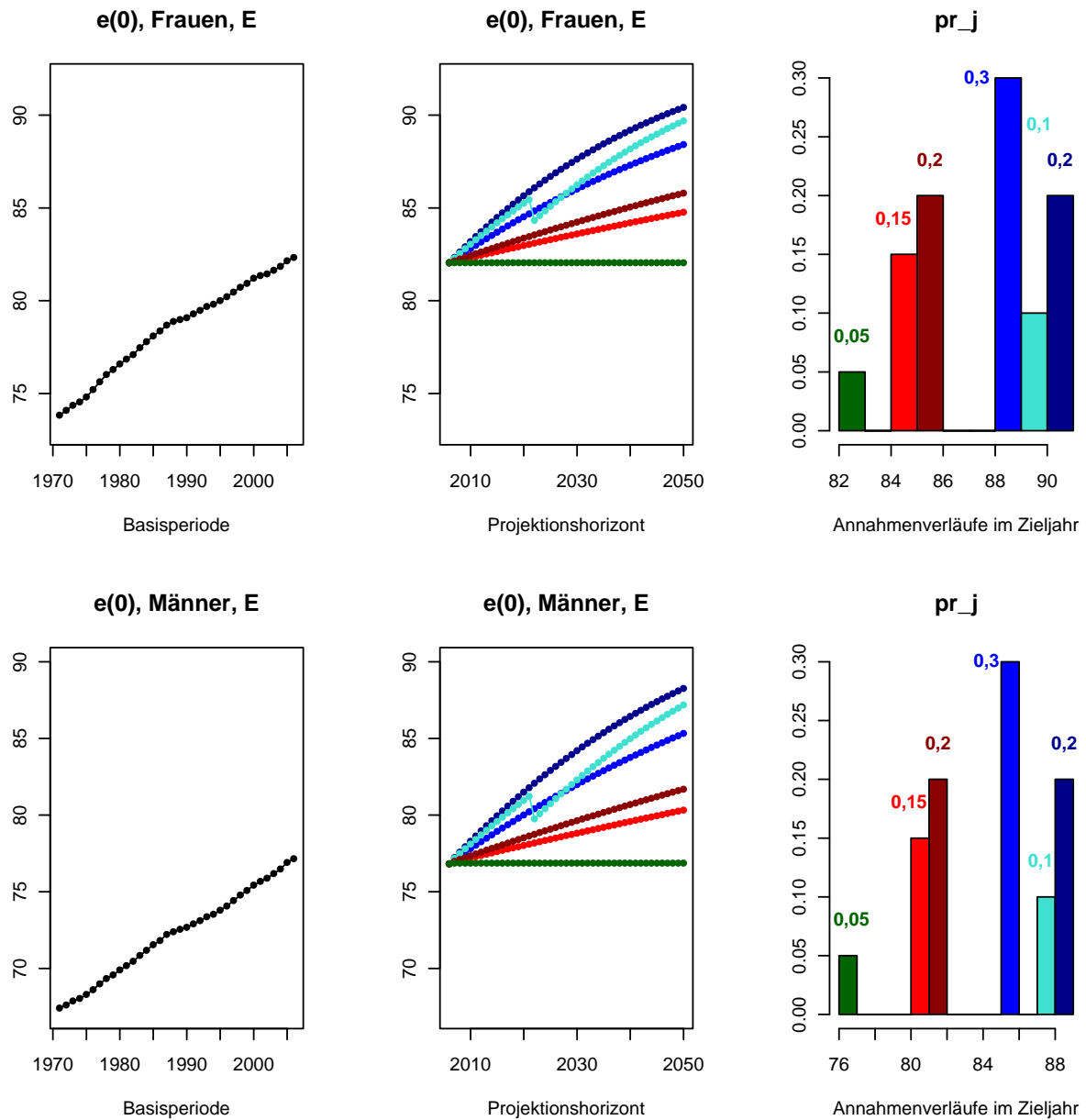


Abbildung 12.2.: Annahmenverläufe  $j$  zur Sterblichkeit (bzw. zur *Lebenserwartung bei Geburt*,  $e(0)$ ) für die einheimischen Frauen (oben) und Männer (unten) mit den subjektiv zugeordneten Eintrittswahrscheinlichkeiten  $pr_j$

wird durch die drei blauen Annahmenverläufe repräsentiert. Der türkise Annahmenverlauf unterbricht den linearen Anstieg, um zu zeigen, dass im PPPM auch Annahmenverläufe eingegeben werden können, die sich im Niveau und im Verlauf unterscheiden. So könnte ein abruptes Absinken der Lebenserwartung bei Geburt z. B. durch eine Epidemie hervorgerufen

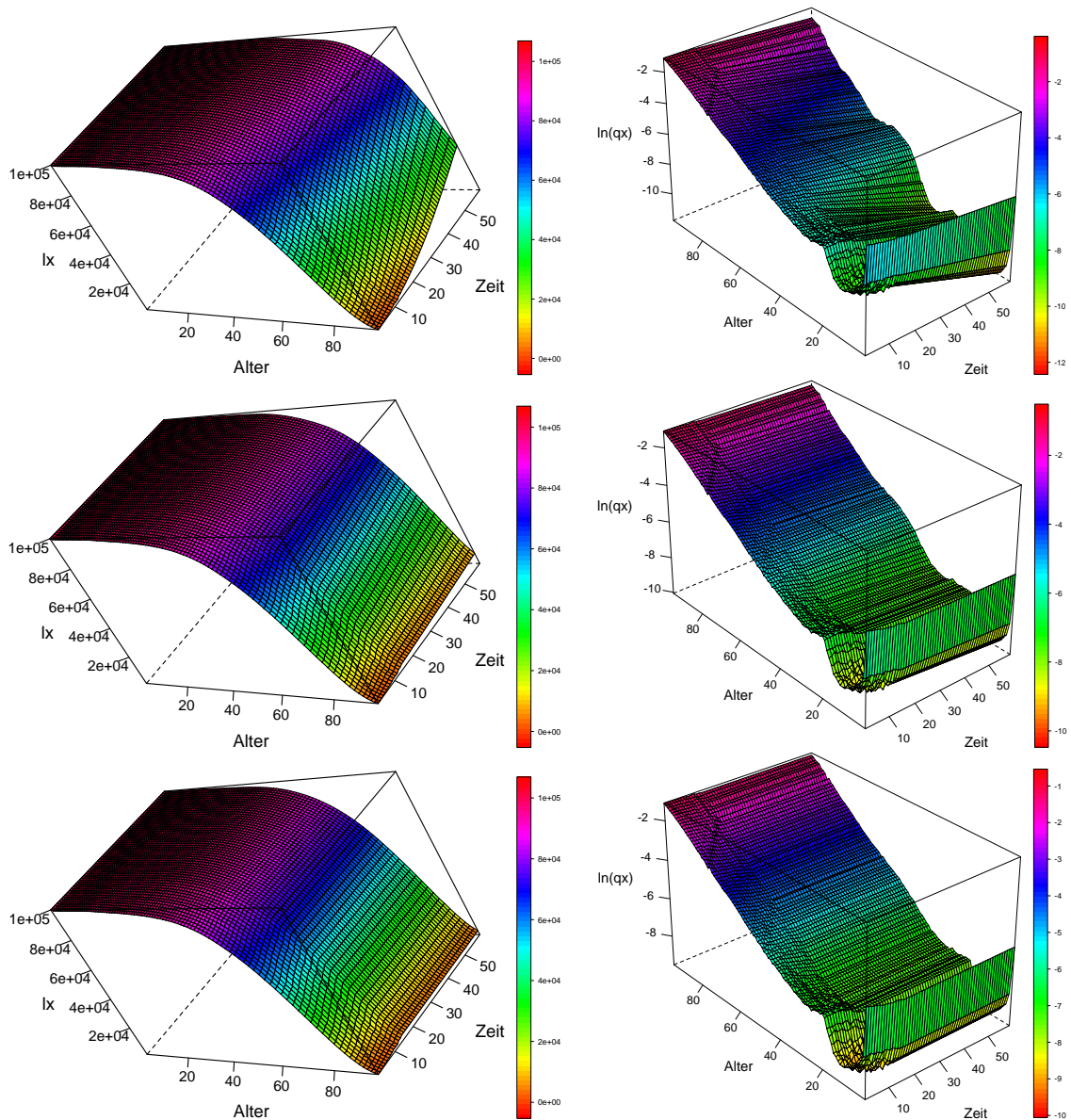


Abbildung 12.3.: Von oben nach unten: Dunkelblauer, roter und grüner Annahmenverlauf des Modellparameters *Überlebende im Alter  $x$*  (links) bzw. der entsprechenden logarithmierten altersspezifischen Sterbewahrscheinlichkeiten (rechts) der einheimischen Männer über die Basisperiode von 1992 bis 2006 und den Projektionshorizont von 2007 bis 2050.

werden. In der Summe erhalten die drei blauen Annahmenverläufe die höchste Eintrittswahrscheinlichkeit, da sich bislang keine verlangsamte Sterblichkeitsreduktion andeutet und auch in der Vergangenheit oftmals ultimative Niveaus vorausgesagt, aber immer schon

kurze Zeit später durchbrochen worden sind. Darüber hinaus kann der Langzeittrend in der Zukunft auch abgeschwächt werden, da die medizinischen Eingriffe in den menschlichen Alterungsprozess begrenzt sind. Die Abschwächung des Langzeittrends in der Zukunft wird durch die beiden roten Annahmenverläufe repräsentiert, die in der Summe die zweithöchste Eintrittswahrscheinlichkeit erhalten. Daneben gibt es noch einen weiteren, grünen Annahmenverlauf, der keine Sterblichkeitsveränderungen infolge einer neuen Krankheit oder Ausbreitung gesundheitsgefährdender Risikofaktoren (wie z.B. Adipositas) unterstellt. Der grüne Annahmenverlauf erhält die niedrigste Eintrittswahrscheinlichkeit, da mit ihm lediglich gezeigt werden soll, dass bei einer Prognose mit dem PPPM auch ungewöhnliche (, aber eventuell auch nicht ganz auszuschließende) Annahmenverläufe berücksichtigt werden können.

Das Statistische Bundesamt prognostiziert die Bevölkerung Deutschlands mit der Szenario-Technik (vgl. Abschnitt 3.2.2 ab Seite 41), sodass zwar mehrere Annahmenverläufe für einen Modellparameter berücksichtigt, aber nicht mit einer Eintrittswahrscheinlichkeit gewichtet werden können. In der 12. koordinierten Bevölkerungsprognose [266] werden lediglich zwei Annahmen zur zukünftigen Entwicklung der Sterblichkeit berücksichtigt. Die Basisannahme unterstellt eine mäßige Sterblichkeitsreduktion, die sich aus der Verbindung der kurz- und langfristigen Sterblichkeitsentwicklung ergibt. Der zweite Annahmenverlauf unterstellt hingegen eine hohe Sterblichkeitsreduktion, die sich aus der Fortschreibung des kurzfristigen Trends seit 1970 ergibt. Konkrete Annahmenwerte zur Lebenserwartung bei Geburt im Jahr 2050 liegen bei den Frauen etwa bei 88 und 90 Jahren und bei den Männern bei etwa 83,5 und 86 Jahren. Damit entsprechen die Annahmenverläufe zur Sterblichkeit aus der 12. koordinierten Bevölkerungsprognose den blauen und z. T. roten Annahmenverläufen der Beispielprojektion. Im Vergleich dazu können herkömmliche probabilistische Bevölkerungsprognosemodelle zwar mehrere Annahmenverläufe gewichtet berücksichtigen, allerdings sind diese häufig durch die Vorgabe einer modell-basierten Methode und einer festen Verteilungsform für die Eintrittswahrscheinlichkeiten eingeschränkt (vgl. Abschnitt 4.2). So kann mit diesen Ansätzen zwar eine Annahmenverteilung zur Sterblichkeit konstruiert werden, bei der die verschiedenen Annahmenverläufe (aus zufällig normalverteilten Abweichungen) um einen fest vorgegebenen mittleren Annahmenverlauf streuen. Wenn nun aber ein Prognostiker (wie in der Beispielprojektion) mehrere Annahmenverläufe berücksichtigen möchte, die sich im Niveau, im Verlauf und in der erwarteten Eintrittswahrscheinlichkeit unterscheiden, muss er oder sie auf das PPPM bzw. P3J zurückgreifen, da nur in diesem Modell Annahmenverteilungen derart



flexibel generiert werden können.

**Migranten** In der Beispielprojektion soll ein sukzessiver Anpassungsprozess zwischen den Immigranten und Einheimischen modelliert werden, bei dem der Sterblichkeitsvorteil der direkten Zuzüge schrittweise über die einzelnen Nachkommengenerationen abnimmt, bis sich die Sterblichkeit zwischen Immigranten und Einheimischen nicht mehr unterscheidet.<sup>3</sup>

Zur Modellierung des sukzessiven Anpassungsprozesses zwischen der Sterblichkeit der direkten Zuzüge und der Einheimischen über die drei Nachkommengenerationen werden für die Frauen und Männer jeweils fünf weitere Annahmenverläufe (auf der Basis der geschätzten Jump-off Werte) für die direkten Zuzüge wie auch für die erste Nachkommengeneration extra generiert, wohingegen der zweiten und dritten Nachkommengeneration dieselben Annahmenverläufe wie für die Einheimischen unterstellt werden. Zusammen mit der variablen Zuordnung der Eintrittswahrscheinlichkeiten kann dann ein *gleichmäßiger* Übergang von der Sterblichkeit der direkten Zuzüge zur Sterblichkeit der Einheimischen über die Nachkommengenerationen modelliert werden.

Die Abbildungen 12.4 und 12.5 geben jeweils einen Überblick über die verschiedenen Annahmenverläufe zur Sterblichkeit der Immigrantinnen und Immigranten und deren subjektiv zuzuordnende Eintrittswahrscheinlichkeit. Die Annahmenwerte der *Lebenserwartung bei Geburt* im Zieljahr können zusammen mit den entsprechenden Eintrittswahrscheinlichkeiten für die Immigrantinnen der Tabelle 12.2 und für die Immigranten der Tabelle 12.3 entnommen werden.

Anhand der konkreten Annahmenwerte für die *Lebenserwartung bei Geburt* (im Zieljahr) und der entsprechenden Eintrittswahrscheinlichkeiten kann der *gleichmäßige* Übergang in der Sterblichkeit zwischen den direkten Zuzügen und den Einheimischen (über die einzelnen Nachkommengenerationen) nochmal verdeutlicht werden. So sinkt beispielsweise der jeweils höchste Annahmenwert für die (weibliche) Lebenserwartung bei Geburt (im Zieljahr) von 92,63 bei den direkten Zuzügen über 91,61 bei der ersten Nachkommengeneration auf 90,41 bei der zweiten und dritten Nachkommengeneration, wobei die Eintrittswahrscheinlichkeit für diesen Annahmenwert bei der zweiten Nachkommengeneration mit 0,25 noch etwas

---

<sup>3</sup>Vereinfachend sollen sich die Annahmen zur Sterblichkeit nicht zwischen den Einheimischen und Emigranten unterscheiden.

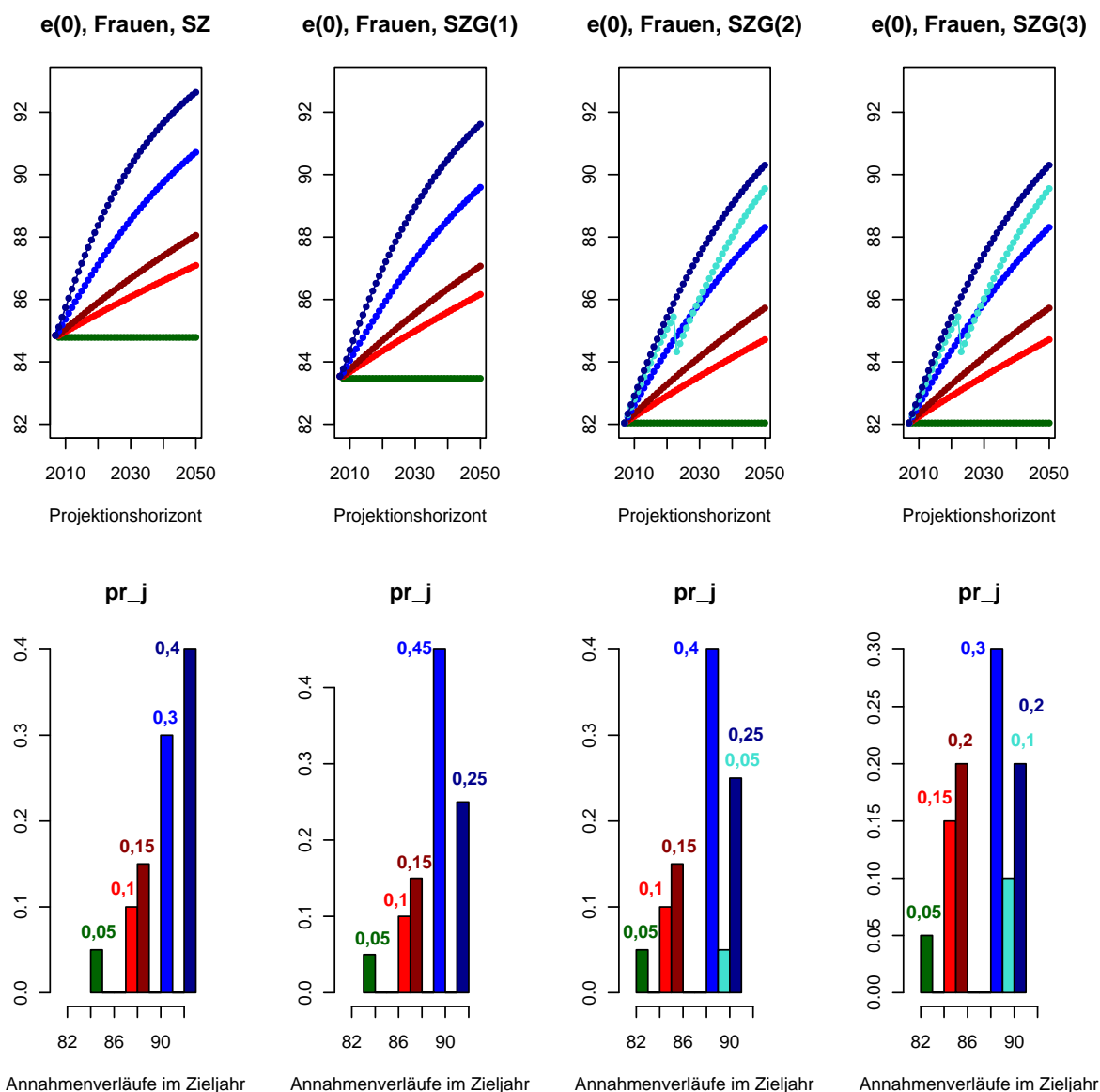


Abbildung 12.4.: Annahmenverläufe  $j$  zur Sterblichkeit (bzw. zur *Lebenserwartung bei Geburt*,  $e(0)$ ) für die Immigrantinnen mit den subjektiv zugeordneten Eintrittswahrscheinlichkeiten  $pr_j$

höher ist als bei der dritten Nachkommengeneration mit 0,2. Ein ähnliches Bild zeigt sich bei den Männern: Auch hier sinkt der jeweils höchste Annahmenwert für die Lebenserwartung bei Geburt (im Zieljahr) von 91,68 bei den direkten Zuzügen über 90,2 bei der ersten Nachkommengeneration auf 88,26 bei der zweiten und dritten Nachkommengeneration, wobei wieder die Eintrittswahrscheinlichkeit für diesen Annahmenwert bei der zweiten Nach-

## 12. Beispielhafte Anwendung des PPPM und P3J

Annahmenverlauf $j$	$e_0^{SZ}$	$pr_j^{SZ}$	$e_0^{SZG(1)}$	$pr_j^{SZG(1)}$	$e_0^{SZG(2)}$	$pr_j^{SZG(2)}$	$e_0^{SZG(3)}$	$pr_j^{SZG(3)}$
Dunkelblau	92,63	0,4	91,61	0,25	90,41*	0,25	90,41*	0,2
Türkis	-	-	-	-	89,68*	0,05	89,68*	0,1
Blau	90,71	0,3	89,59	0,45	88,41*	0,4	88,41*	0,3
Dunkelrot	88,05	0,15	87,07	0,15	85,79*	0,15	85,79*	0,2
Rot	87,09	0,1	86,16	0,1	84,76*	0,1	84,76*	0,15
Grün	84,78	0,05	83,47	0,05	82,03*	0,05	82,03*	0,05

Tabelle 12.2.: Annahmenverläufe  $j$  zur Mortalität der Immigrantinnen charakterisiert durch die *Lebenserwartung bei Geburt*  $e_0$  im Zieljahr und die subjektiv zugeordnete Eintrittswahrscheinlichkeit  $pr_j$ ; die mit einem \* markierten Annahmenverläufe entsprechen jeweils einem Annahmenverlauf der einheimischen Frauen.

Annahmenverlauf $j$	$e_0^{SZ}$	$pr_j^{SZ}$	$e_0^{SZG(1)}$	$pr_j^{SZG(1)}$	$e_0^{SZG(2)}$	$pr_j^{SZG(2)}$	$e_0^{SZG(3)}$	$pr_j^{SZG(3)}$
Dunkelblau	91,68	0,4	90,2	0,25	88,26*	0,25	88,26*	0,2
Türkis	-	-	-	-	87,17*	0,05	87,17*	0,1
Blau	88,91	0,3	87,22	0,45	85,32*	0,4	85,32*	0,3
Dunkelrot	85,2	0,15	83,37	0,15	81,68*	0,15	81,68*	0,2
Rot	83,91	0,1	82,27	0,1	80,32*	0,1	80,32*	0,15
Grün	80,88	0,05	78,89	0,05	76,86*	0,05	76,86*	0,05

Tabelle 12.3.: Annahmenverläufe  $j$  zur Mortalität der Immigranten charakterisiert durch die *Lebenserwartung bei Geburt*  $e_0$  im Zieljahr und die subjektiv zugeordnete Eintrittswahrscheinlichkeit  $pr_j$ ; die mit einem \* markierten Annahmenverläufe entsprechen jeweils einem Annahmenverlauf der einheimischen Männer.

kommengeneration mit 0,25 noch etwas höher ist als bei der dritten Nachkommengeneration mit 0,2. Demzufolge nimmt der Sterblichkeitsvorteil der direkten Zuzüge gegenüber den Einheimischen sowohl bei den Frauen als auch bei den Männern sukzessive über die einzelnen Nachkommengenerationen ab, wobei die Sterblichkeit der Frauen immer etwas niedriger ist als die der Männer.

Im Gegensatz zum PPPM bzw. P3J können in der 12. koordinierten Bevölkerungsprognose vom Statistischen Bundesamt (wie auch bei herkömmlichen probabilistischen Bevölkerungsprognosemodellen) keine separaten Annahmen zur Sterblichkeit für die Einheimischen, die direkten Zu- und Fortzüge und deren Nachkommengenerationen getroffen werden, sodass

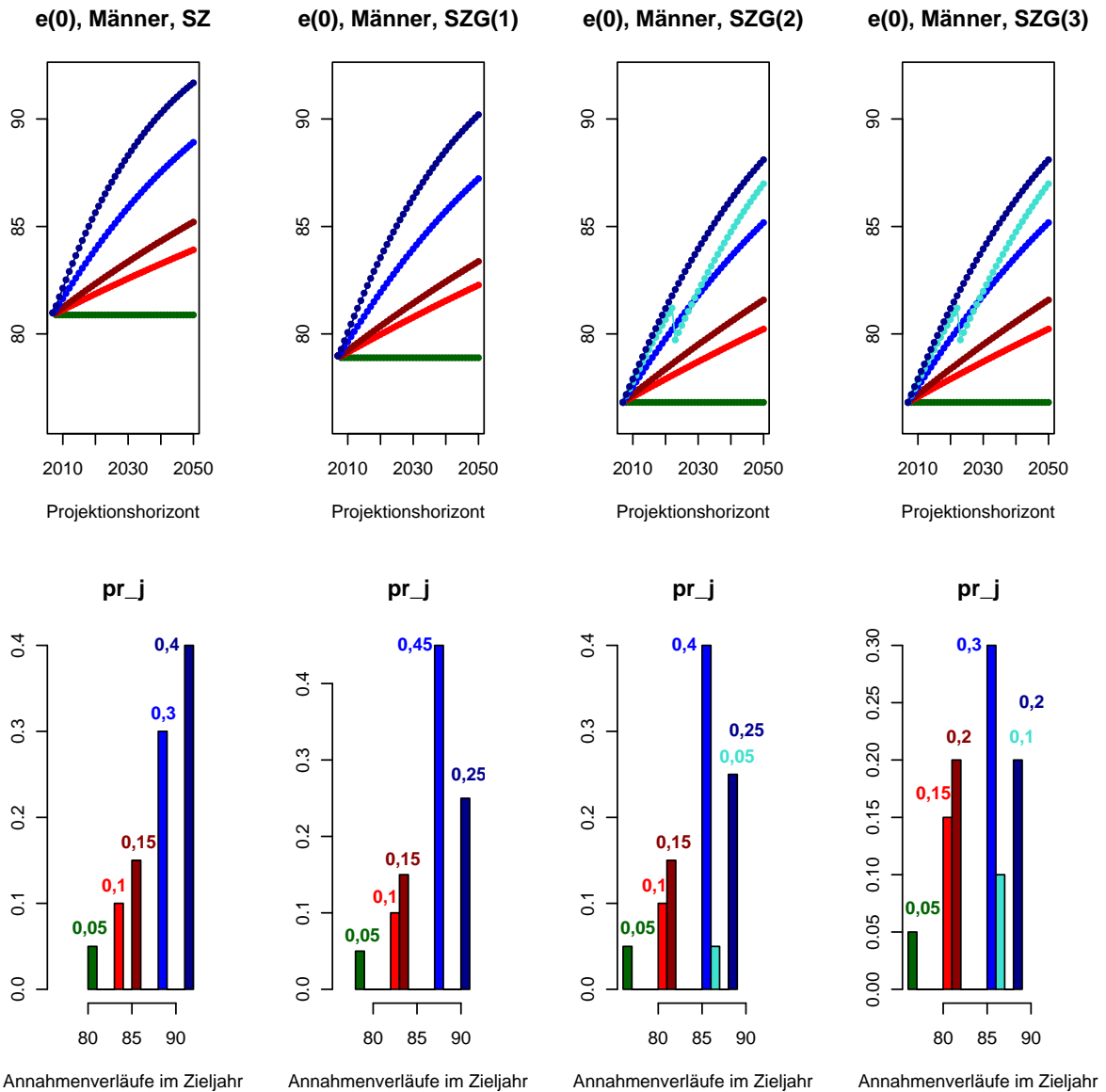


Abbildung 12.5.: Annahmenverläufe  $j$  zur Sterblichkeit (bzw. zur *Lebenserwartung bei Geburt*,  $e(0)$ ) für die Immigranten mit den subjektiv zugeordneten Eintrittswahrscheinlichkeiten  $pr_j$

diese Modelle dieselbe Sterblichkeit für alle Personen (desselben Alters und Geschlechts) unterstellen. Folglich kann (aufgrund der Prognose der einzelnen Subpopulationen) nur mit dem PPPM bzw. P3J ein Anpassungsprozess zwischen der niedrigeren Sterblichkeit der direkten Zuzüge und der höheren Sterblichkeit der Einheimischen über die einzelnen Nachkommengenerationen modelliert werden.

### 12.4.2. Fertilität

#### Methode

Die Annahmenverläufe zur Fertilität, die jeweils die Fertilitätsraten einer Subpopulation über alle (reproduktiven) Alter über den gesamten Projektionshorizont enthalten, werden mit einem Methodenmix erstellt. Dabei wird die Fertilität mit der Modellgleichung des Ansatzes von Lee und Carter (vgl. Abschnitt 5.2.1 ab Seite 84) bzw. Lee (vgl. Abschnitt 6.1.1 ab Seite 117) prognostiziert, wobei die  $\beta_x$  mit Expertenmeinung bestimmt und die  $k_t$  (ausgehend von den mit der Singulärwertzerlegung geschätzten Werten für die Basisperiode) mit einem autoregressiven Modell (der Ordnung 1) und einem subjektiv festzulegenden Drift fortgeschrieben werden.

#### Bestimmung der Jump-off Werte

Um die Annahmenverläufe für den Modellparameter *altersspezifische Fertilitätsraten* zu generieren, müssen zunächst die Werte aus dem Jump-off Jahr für jede Subpopulation ermittelt werden.

Grundsätzlich liegen altersspezifische Fertilitätsraten für die einheimische und ausländische Bevölkerung aus der amtlichen Statistik für das Berichtsjahr 2006 vor. Im Abschnitt 7.1.2 (ab Seite 138) ist aber bereits erläutert worden, dass die Fertilität der ausländischen Bevölkerung nicht mit der Fertilität der direkten Zuzüge und deren erster (und weiterer) Nachkommengeneration(en) übereinstimmt, sodass diese entsprechend angepasst werden muss. Dabei werden für die direkten Zuzüge alle altersspezifischen Fertilitätsraten der Ausländerinnen mit dem Faktor 1,1 multipliziert; dadurch kann sowohl das durchschnittlich jüngere Alter der Mütter bei Geburt als auch die höhere Fertilität der direkten Zuzüge (gegenüber den Einheimischen) angemessen eingefangen werden. Anschließend wird die Fertilität der ersten Nachkommengeneration der direkten Zuzüge als arithmetisches Mittel aus der Fertilität der direkten Zuzüge und der ausländischen Bevölkerung ermittelt (vgl. Abbildung 12.6).

Ab der zweiten Nachkommengeneration werden kaum noch Unterschiede zur Fertilität der Einheimischen erwartet, sodass dieselben altersspezifischen Fertilitätswerte als Jump-off Werte herangezogen werden.

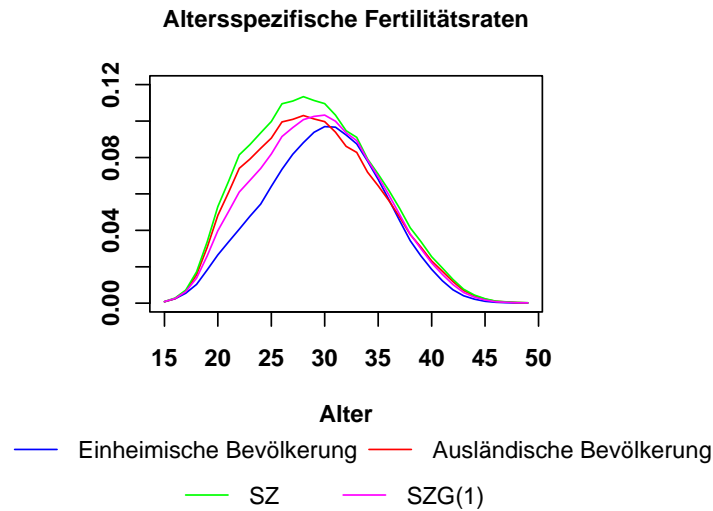


Abbildung 12.6.: Bestimmung der Jump-off Werte für die altersspezifischen Fertilitätsraten der direkten Zuzüge (SZ) und ihrer ersten Nachkommengeneration (SZG(1)) aus den altersspezifischen Fertilitätsraten der einheimischen und ausländischen Bevölkerung des Jahres 2006; Datenquelle: Statistisches Bundesamt

### Vergangene Entwicklung und Variationen des Langzeittrends

**Einheimische** Die Abbildung 12.7 zeigt sowohl die vergangene Entwicklung als auch 4 verschiedene Annahmenverläufe zur zukünftigen Entwicklung der Fertilität mit den subjektiv zugeordneten Eintrittswahrscheinlichkeiten für die einheimischen Frauen anhand des Parameters *Total Fertility Rate*. Zudem enthält die Tabelle 12.4 die konkreten Annahmenwerte zur Total Fertility Rate und dem durchschnittlichen Alter der Mütter bei Geburt im Zieljahr wie auch die entsprechenden Eintrittswahrscheinlichkeiten.

Annahmenverlauf $j$	$TFR$	$\bar{x}_f$	$pr_j$
Dunkelblau	1,59	29,43	0,2
Blau	1,507	32,54	0,1
Rot	1,42	31,17	0,6
Grün	1,31	29,63	0,1

Tabelle 12.4.: Annahmenverläufe  $j$  zur Fertilität der Einheimischen charakterisiert durch die Total Fertility Rate (TFR), das durchschnittliche Alter der Mütter bei Geburt ( $\bar{x}_f$ ) und die subjektiv zugeordnete Eintrittswahrscheinlichkeit  $pr_j$ .

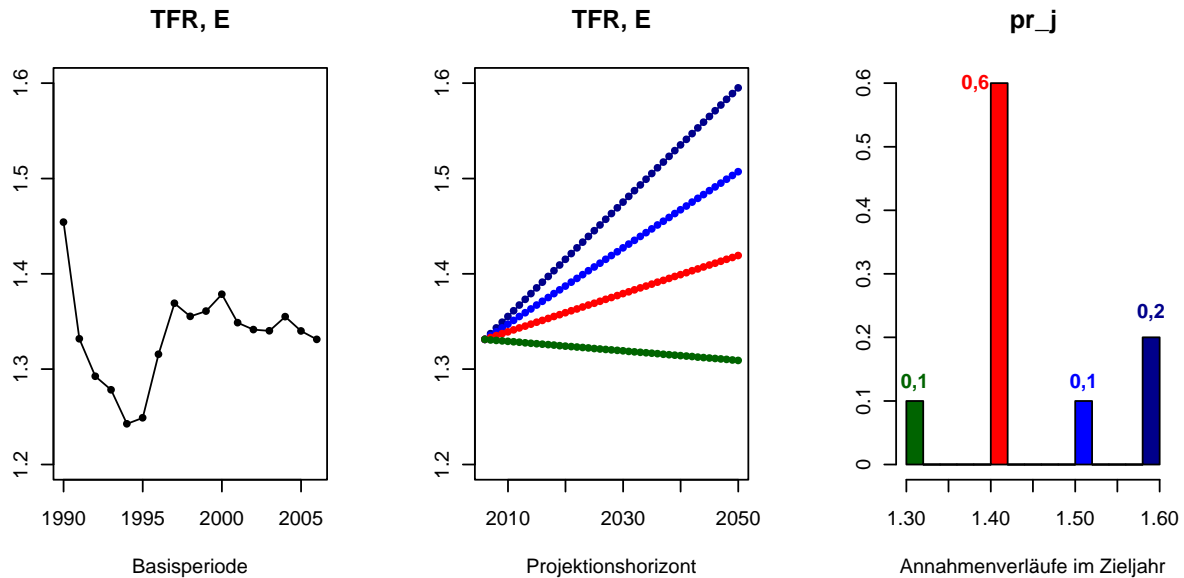


Abbildung 12.7.: Annahmenverläufe  $j$  zur Fertilität (bzw. zur *Total Fertility Rate* (TFR)) für die einheimischen Frauen mit den subjektiv zugeordneten Eintrittswahrscheinlichkeiten  $pr_j$

In der Vergangenheit schwankt die TFR in Deutschland zuletzt um einen Mittelwert von 1,36. Folglich wird der Langzeittrend mit dem grünen und roten Annahmenverlauf fortgeschrieben, wobei der grüne Annahmenverlauf ein weiteres Absinken auf 1,31 und der rote Annahmenverlauf ein leichtes Ansteigen der TFR auf 1,4 (im Zieljahr) annimmt. Darüber hinaus gehen die beiden blauen Annahmenverläufe von einer leichten Erhöhung der TFR auf 1,507 und 1,59 in der Zukunft aus, um einen Wertewandel oder auch etwaige Reaktionen auf pronatalistische familienpolitische Maßnahmen (wie z.B. die Einführung des Elterngelds 2007 [44] oder den Beschluss des Tagesbetreuungsausbaugesetzes von 2005 [43]) einzufangen. Neben dem Niveau unterscheiden sich die 4 Annahmenverläufe zur Fertilität auch im Verlauf (über das Alter und die Zeit). So nehmen zwar sowohl der dunkelblaue als auch der blaue Annahmenverlauf eine Erhöhung der Fertilität an, allerdings unterstellt der blaue Annahmenverlauf ein deutlich höheres Durchschnittsalter der Mütter bei Geburt als der dunkelblaue Annahmenverlauf. Folglich basiert der Fertilitätsanstieg beim dunkelblauen Annahmenverlauf auf einer Erhöhung der Fertilität bei den jungen *und* älteren Müttern, wohingegen er bei dem blauen Annahmenverlauf ausschließlich auf der Erhöhung der Fertilität bei den älteren Müttern basiert. Daneben basieren auch der leichte Fertilitätsanstieg bzw. -rückgang beim roten bzw. grünen Annahmenverlauf auf einer leichten

Zu- bzw. Abnahme der Fertilität bei den älteren bzw. jüngeren Müttern (vgl. Abbildung 12.8).

In der 12. koordinierten Bevölkerungsprognose vom Statistischen Bundesamt [266] werden drei Annahmen zur zukünftigen Entwicklung der Fertilität berücksichtigt. Der Hauptannahmenverlauf unterstellt die annähernde Konstanz des Langzeittrends mit einer TFR von 1,4, wobei das durchschnittliche Alter der Mütter bei Geburt auf 31,4 ansteigt. Daneben gibt es zwei Annahmenverläufe, die einen Anstieg bzw. Rückgang der TFR auf 1,6 bzw. 1,2 unterstellen, wobei das durchschnittliche Alter der Mütter bei Geburt auf 30,9 bzw. 31,9 ansteigt. Folglich stimmen die erwarteten Entwicklungsspektren zur Fertilität von der 12. koordinierten Bevölkerungsprognose und der Beispielprojektion bis auf den Annahmenverlauf mit dem relativ starken Rückgang der TFR auf 1,2, der in der Beispielprojektion ausgeschlossen wird, nahezu überein. Allerdings können (wie bei den Annahmen zur Sterblichkeit) nur mit dem PPPM bzw. P3J Annahmenverläufe, die verschiedene Zukunftserwartungen repräsentieren, variabel gewichtet in einer Bevölkerungsprognose berücksichtigt werden.

**Migranten** In der Beispielprojektion soll (wie bei den Annahmen zur Sterblichkeit) ein sukzessiver Anpassungsprozess von der höheren Fertilität der direkten Zuzüge zur niedrigeren Fertilität der Einheimischen über die einzelnen Nachkommengenerationen modelliert werden.<sup>4</sup>

Zur Modellierung des sukzessiven Anpassungsprozesses zwischen der Fertilität der direkten Zuzüge und der Einheimischen über die drei Nachkommengenerationen werden jeweils vier weitere Annahmenverläufe (auf der Basis der geschätzten Jump-off Werte) für die direkten Zuzüge und die erste Nachkommengeneration und ein weiterer Annahmenverlauf für die zweite Nachkommengeneration extra generiert, wohingegen die restlichen Annahmenverläufe der zweiten und dritten Nachkommengeneration mit denen der Einheimischen übereinstimmen. Die Verbindung dieser Annahmenverläufe mit den subjektiv zuzuordnenden Eintrittswahrscheinlichkeiten bewirkt einen *gleichmäßigen* Übergang in der Fertilität zwischen den direkten Zuzügen und Einheimischen.

Die Abbildung 12.9 zeigt die verschiedenen Annahmenverläufe zur Fertilität der Immigrantinnen zusammen mit deren subjektiv zuzuordnender Eintrittswahrscheinlichkeit. Darüber hinaus sind die konkreten Annahmenwerte der TFR im Zieljahr mit den jeweiligen

---

<sup>4</sup>Vereinfachend sollen sich die Annahmen zur Fertilität nicht zwischen den Einheimischen und Emigranten unterscheiden.



## 12. Beispielhafte Anwendung des PPPM und P3J

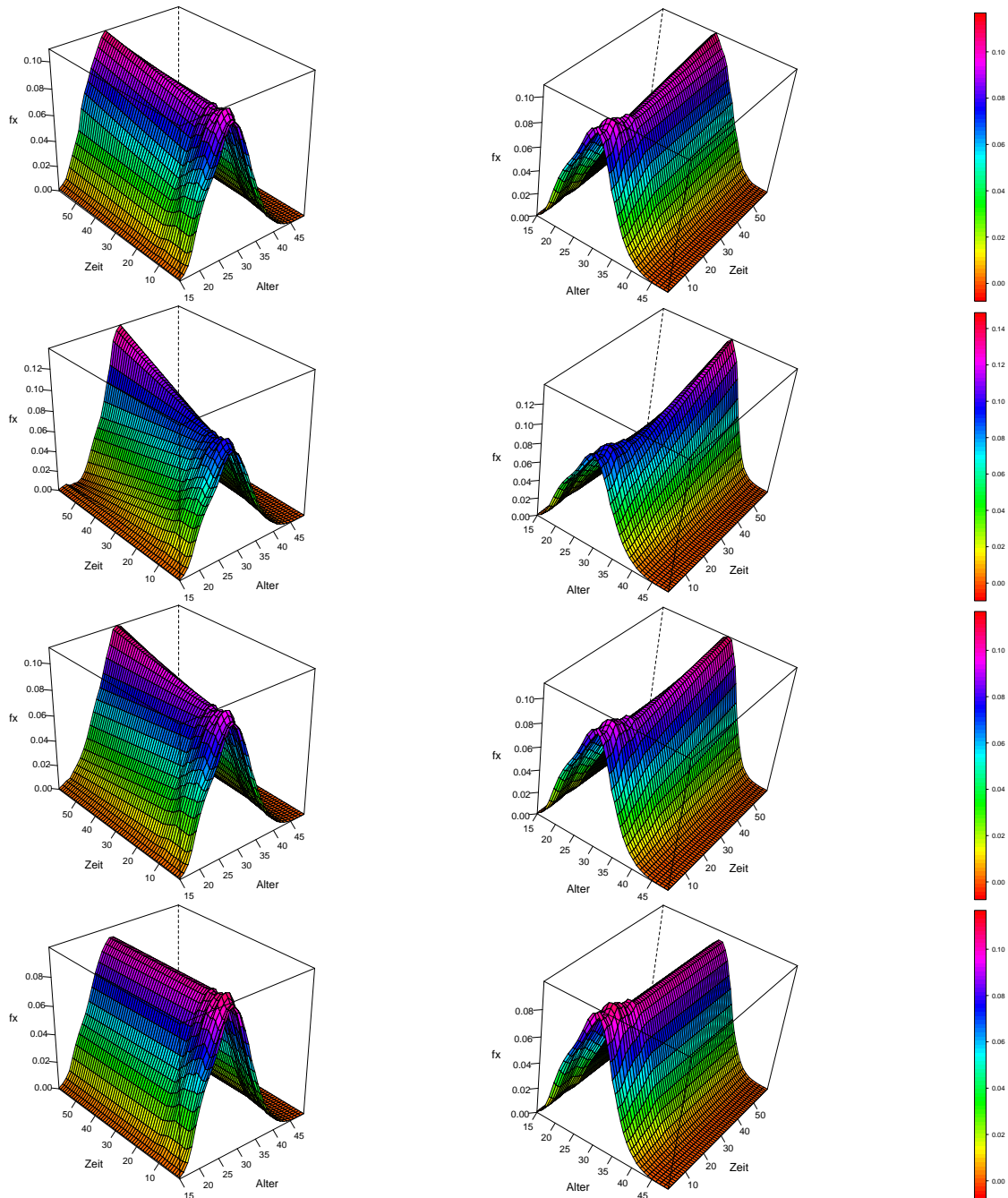


Abbildung 12.8.: Von oben nach unten: Dunkelblauer, blauer, roter und grüner Annahmenverlauf des Modellparameters *altersspezifische Fertilitätsraten* der einheimischen Frauen über die Basisperiode von 1992 bis 2006 und den Projektionshorizont von 2007 bis 2050.

Eintrittswahrscheinlichkeiten in der Tabelle 12.5 für die Immigrantinnen aufgeführt.

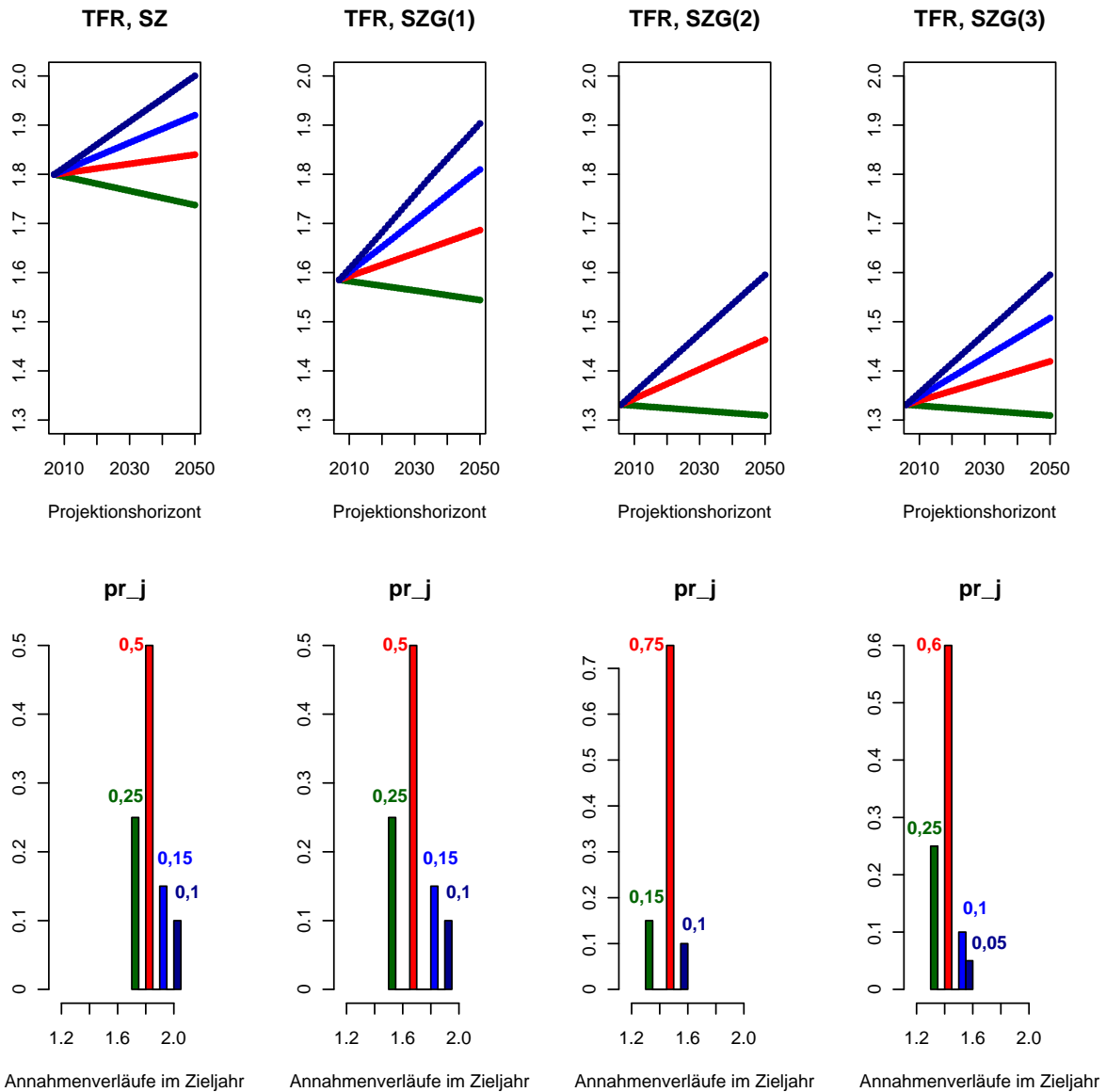


Abbildung 12.9.: Annahmenverläufe  $j$  zur Fertilität (bzw. zur *Total Fertility Rate* (TFR)) für die Immigrantinnen mit den subjektiv zugeordneten Eintrittswahrscheinlichkeiten  $pr_j$

Anhand der konkreten Annahmenwerte für die TFR (im Zieljahr) und der entsprechenden Eintrittswahrscheinlichkeiten kann der *gleichmäßige* Übergang in der Fertilität zwischen den direkten Zuzügen und den Einheimischen (über die einzelnen Nachkommengenerationen)

## 12. Beispielhafte Anwendung des PPPM und P3J

Annahmenverlauf $j$	$TFR^{SZ}$	$pr_j^{SZ}$	$TFR^{SZG(1)}$	$pr_j^{SZG(1)}$	$TFR^{SZG(2)}$	$pr_j^{SZG(2)}$	$TFR^{SZG(3)}$	$pr_j^{SZG(3)}$
Dunkelblau	2	0,1	1,9	0,1	1,59*	0,1	1,59*	0,05
Blau	1,92	0,15	1,81	0,15	-	-	1,507*	0,1
Rot	1,84	0,5	1,69	0,5	1,46	0,75	1,419*	0,6
Grün	1,74	0,25	1,54	0,25	1,31*	0,15	1,309*	0,25

Tabelle 12.5.: Annahmenverläufe  $j$  zur Fertilität der Immigrantinnen charakterisiert durch die Total Fertility Rate (TFR) und die subjektiv zugeordnete Eintrittswahrscheinlichkeit  $pr_j$ ; die mit einem \* markierten Annahmenverläufe entsprechen jeweils einem Annahmenverlauf der einheimischen Frauen.

nochmal verdeutlicht werden. So ist bei gleicher Eintrittswahrscheinlichkeit von 0,1 der höchste TFR-Wert der direkten Zuzüge mit 2 größer als bei der ersten Nachkommengeneration mit 1,9. Außerdem haben zwar die zweite und dritte Nachkommengeneration denselben höchsten TFR-Wert von 1,59, allerdings unterscheiden sie sich in der Höhe der Eintrittswahrscheinlichkeit; so ist die Eintrittswahrscheinlichkeit bei der zweiten Nachkommengeneration mit 0,1 höher als bei der dritten Nachkommengeneration mit 0,05.

Im Gegensatz zum PPPM bzw. P3J können in der 12. koordinierten Bevölkerungsprognose vom Statistischen Bundesamt (wie auch bei herkömmlichen probabilistischen Bevölkerungsprognosemodellen) keine separaten Annahmen zur Fertilität für die Einheimischen, die direkten Zu- und Fortzüge und deren Nachkommengenerationen getroffen werden, sodass dieselbe Fertilität für alle Frauen desselben Alters unterstellt wird. Folglich kann (aufgrund der Prognose der einzelnen Subpopulationen) nur mit dem PPPM bzw. P3J ein Anpassungsprozess von der höheren Fertilität der direkten Zuzüge an die niedrigere Fertilität der Einheimischen über die einzelnen Nachkommengenerationen modelliert werden.

### 12.4.3. Migration

#### Methode

Die verschiedenen Annahmenverläufe zur Migration, die jeweils die Anzahl der direkten Zu- und Fortzüge über alle Alter über den gesamten Projektionshorizont enthalten, werden mit einem Methodenmix erstellt. Dabei werden Veränderungen in der Altersstruktur anhand der vergangenen Entwicklung und bestimmter Zukunftserwartungen subjektiv bestimmt und anschließend mit einem autoregressives Modell (der Ordnung 1) und einem abermals subjektiv zu bestimmenden Drift fortgeschrieben.

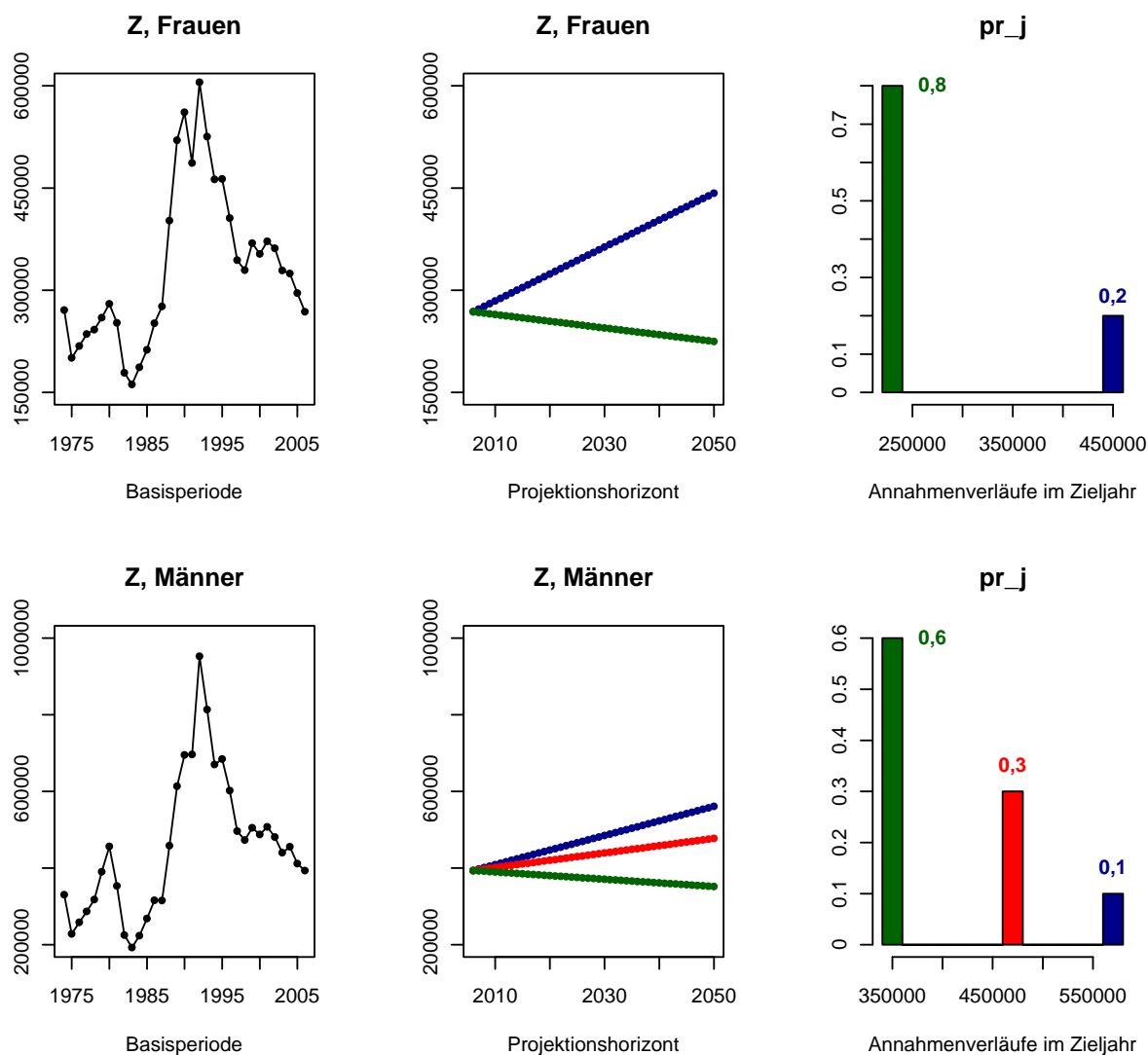


Abbildung 12.10.: Annahmenverläufe  $j$  für die weiblichen (oben) und männlichen (unten) jährlichen Zuzüge ( $Z$ ) mit den subjektiv zugeordneten Eintrittswahrscheinlichkeiten  $pr_j$

### Vergangene Entwicklung und Variationen des Langzeittrends

Die Abbildungen 12.10 und 12.11 zeigen neben der vergangenen Entwicklung auch die Annahmenverläufe für die direkten Zu- und Fortzüge nach Geschlecht und deren subjektiv zugeordnete Eintrittswahrscheinlichkeiten. Die konkreten Annahmenwerte der direkten Zu- und Fortzüge für die Frauen und Männer im Zieljahr 2050 sind zudem in der Tabelle 12.6 aufgeführt.

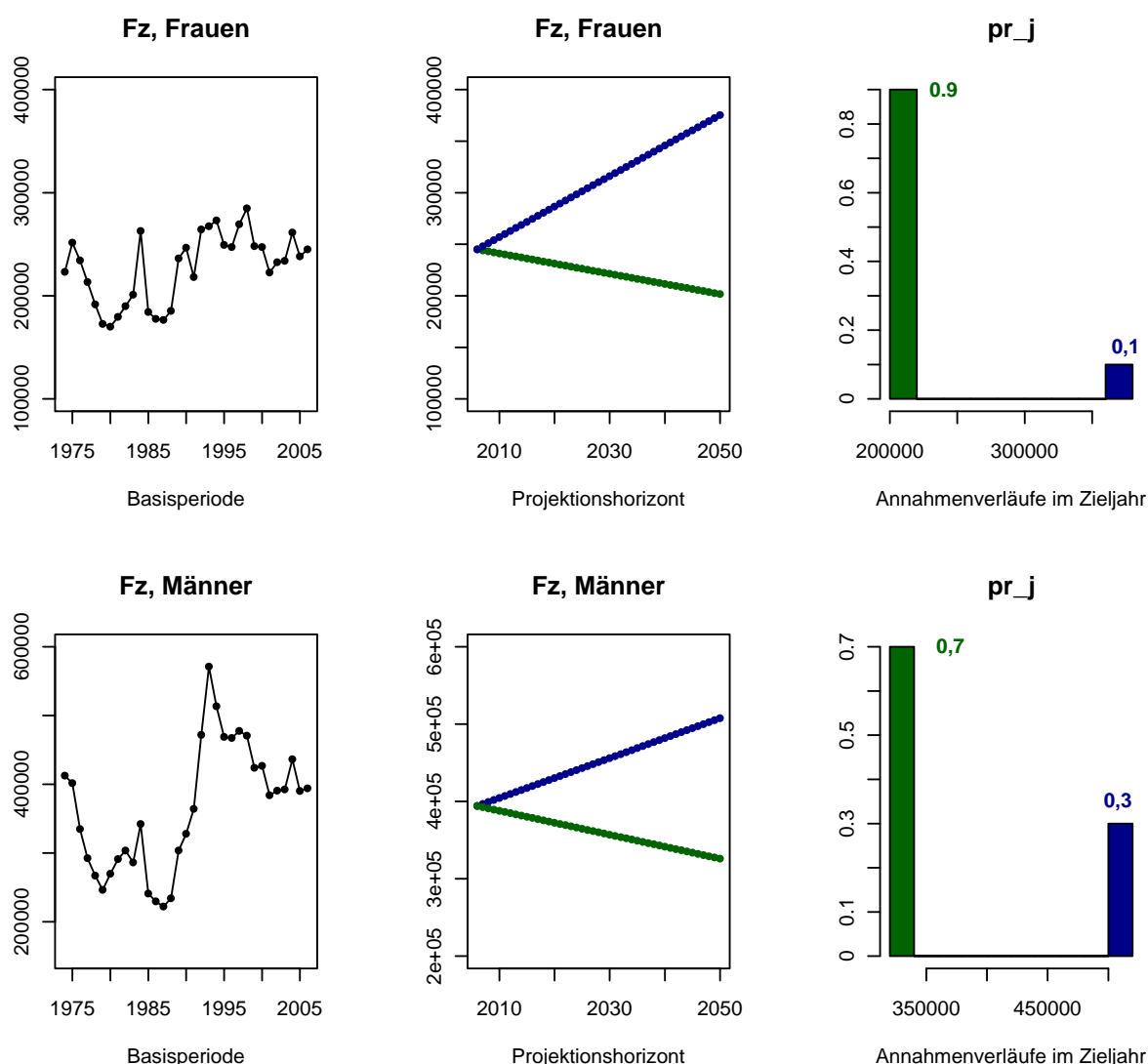


Abbildung 12.11.: Annahmenverläufe  $j$  für die weiblichen (oben) und männlichen (unten) jährlichen Fortzüge (Fz) mit den subjektiv zugeordneten Eintrittswahrscheinlichkeiten  $pr_j$

In der Vergangenheit unterliegt die Zahl der Fortzüge und ganz besonders die der Zuzüge relativ großen Schwankungen, die sich aus dem komplexen, kaum vorhersehbaren Zusammenspiel sozialer, ökonomischer und politisch(-rechtlich)er Bedingungen in den Ziel- und Herkunftsländern ergeben. Aufgrund dieser Unsicherheit soll die Zahl der direkten Zu- und Fortzüge (in einem vorgegebenen Rahmen) in der Beispielprojektion relativ stark schwanken können, sodass immer mindestens zwei Annahmenverläufe, die jeweils einen

Annahmenverlauf $j$	Frauen				Männer			
	$Z$	$pr_j^Z$	$Fz$	$pr_j^{Fz}$	$Z$	$pr_j^Z$	$Fz$	$pr_j^{Fz}$
Dunkelblau	442.380	0,2	375.238	0,1	560.870	0,1	507.567	0,3
Rot	-	-	-	-	477.226	0,3	-	-
Grün	224.751	0,8	201.577	0,9	351.783	0,6	325.975	0,7

Tabelle 12.6.: Annahmenverläufe  $j$  zur Bruttomigration charakterisiert durch die weiblichen und männlichen Zu- und Fortzüge im Zieljahr und die subjektiv zugeordnete Eintrittswahrscheinlichkeit  $pr_j$ .

Rückgang (grün) und einen Anstieg (dunkelblau) unterstellen, generiert werden. Darüber hinaus weist die Altersstruktur der Zu- und Fortzüge sowohl für die Männer als auch für die Frauen eine regelmäßige Struktur auf (vgl. Abbildung 12.12), die bei der Generierung der einzelnen Annahmenverläufe in ihrer groben Form einigermaßen erhalten bleiben soll. Daher werden die größten Veränderungen in den Altersbereichen angenommen, in denen ohnehin das größte Wanderungsgeschehen stattfindet.

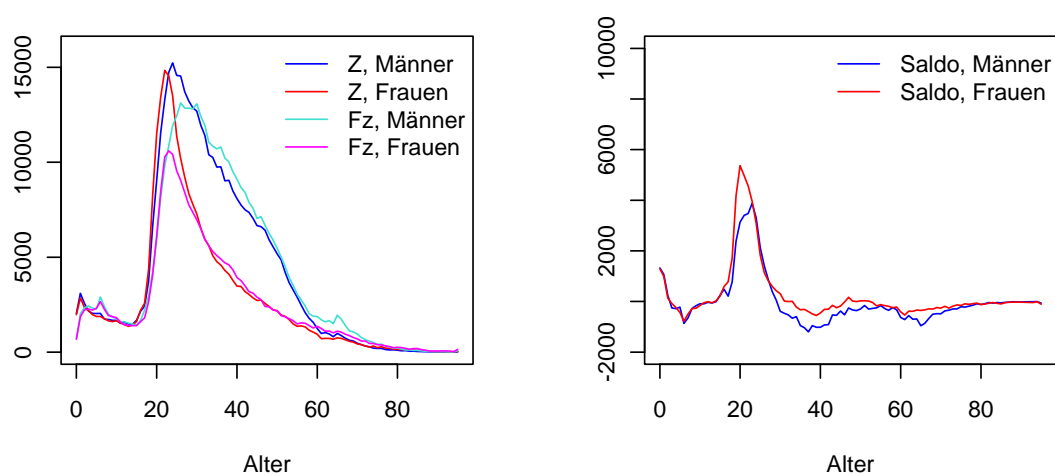


Abbildung 12.12.: Alters- und Geschlechtsstruktur der Zu- und Fortzüge im Jahr 2006; Datenquelle: Statistisches Bundesamt

Zunächst sollen exemplarisch die grünen Annahmenverläufe der Frauen, für die die höchsten Eintrittswahrscheinlichkeiten angenommen werden, im Detail über das Alter und die Zeit dargestellt und erläutert werden (vgl. Abbildung 12.13). Sowohl bei den Zu- als auch bei den Fortzügen wird ein Rückgang angenommen, allerdings findet dieser Rückgang bei

den Zuzügen vornehmlich bei den jungen Erwachsenen im Alter zwischen 20 und 25 statt, wohingegen er bei den Fortzügen im jungen und mittleren Erwachsenenalter von 20 bis 50 stattfindet.

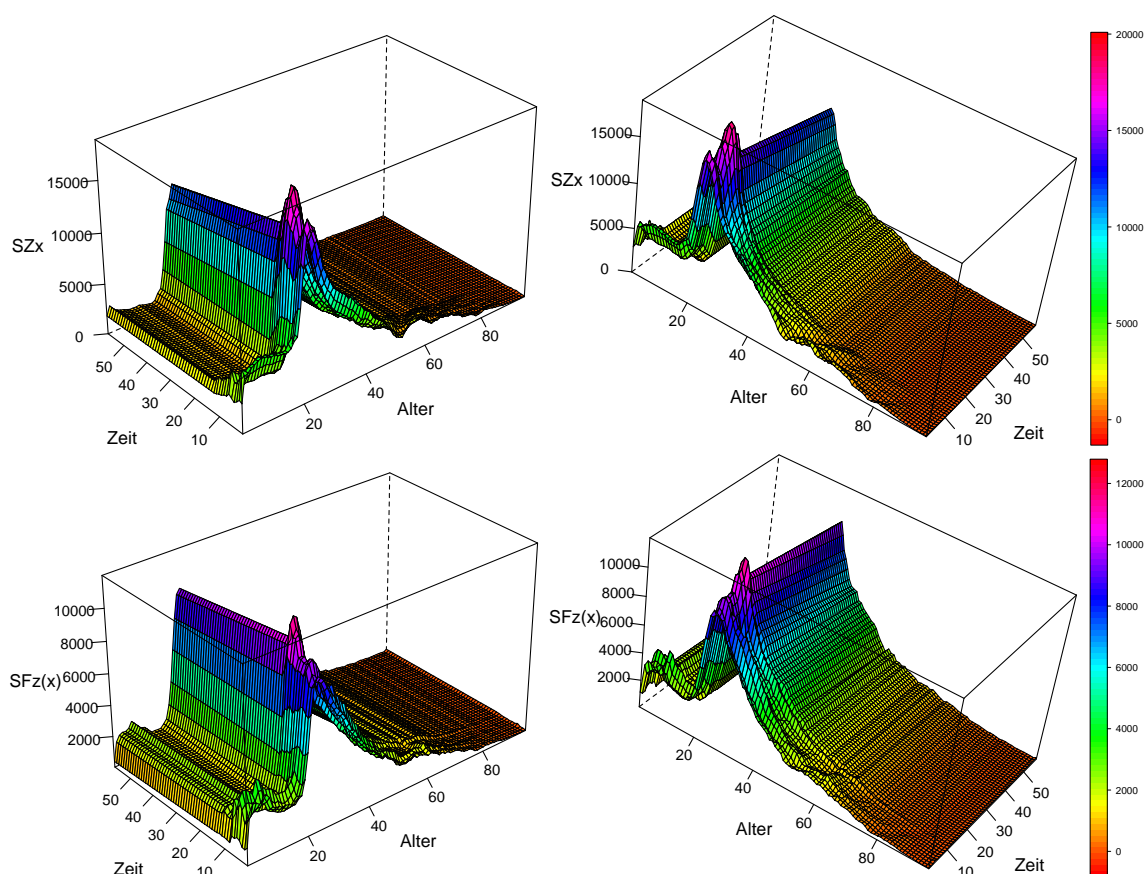


Abbildung 12.13.: Grüne Annahmenverläufe der jährlichen Zuzüge (oben) und Fortzüge (unten) nach dem Alter und der Zeit von 1994 bis 2050 für die Frauen

Aus der Kombination der grünen Annahmenverläufe für die direkten Zu- und Fortzüge der Frauen resultiert ein positiver Wanderungssaldo, da für jedes Projektionsintervall etwas mehr Zu- als Fortzüge angenommen werden. Im Gegensatz dazu zeigt nun die Abbildung 12.14 exemplarisch den grünen und dunkelblauen Annahmenverlauf der direkten Zu- und Fortzüge der Männer, aus deren Kombination ein negativer Wanderungssaldo resultiert, da für jedes Projektionsintervall mehr Fort- als Zuzüge angenommen werden. Außerdem ist zu erkennen, dass für alle Alter der direkten Zuzüge ein Rückgang angenommen wird, wobei der größte Rückgang bei den jungen und mittleren Erwachsenenaltern, also von 20 bis 35, angenommen

wird. Ein umgekehrtes Bild zeigt sich hingegen bei den direkten Fortzügen: Anstatt eines Rückgangs wird hier für alle Alter ein Anstieg angenommen, wobei der größte Anstieg über die jungen, mittleren und hohen Erwachsenenalter, also von 20 bis 70, angenommen wird.

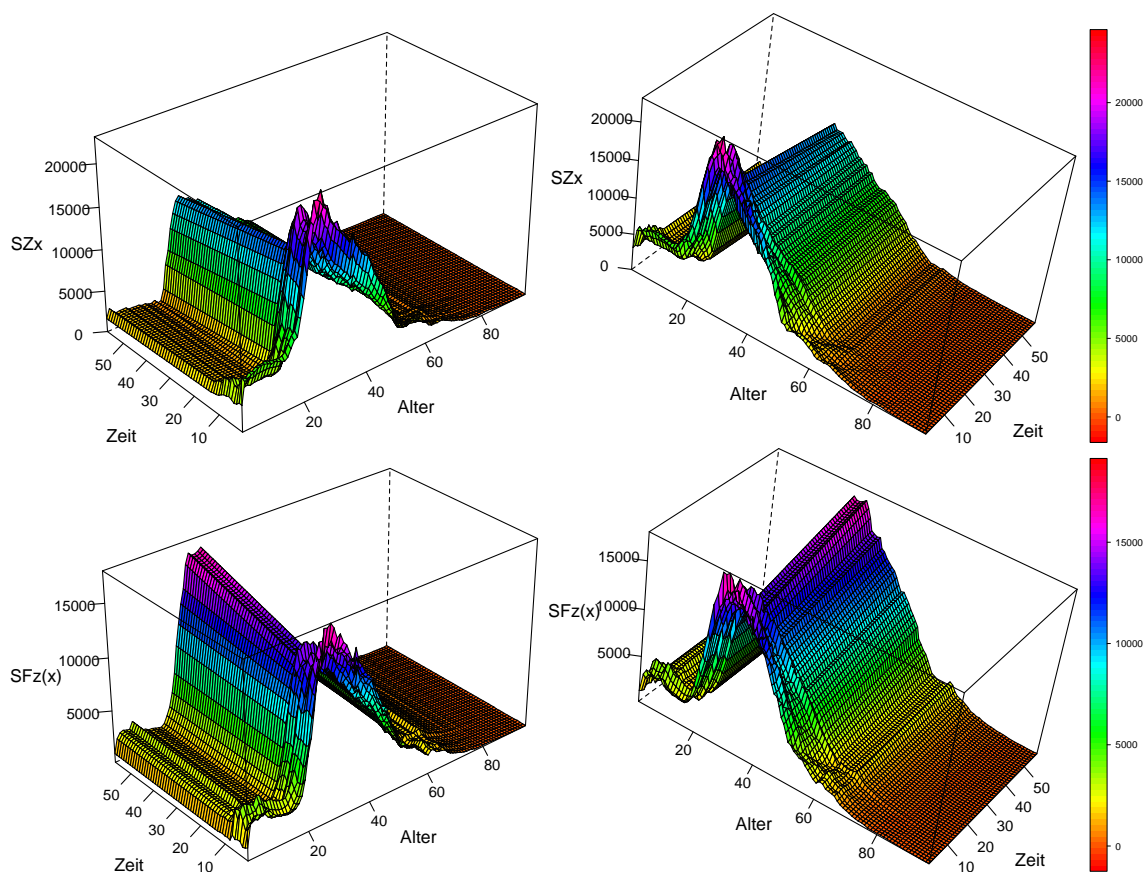


Abbildung 12.14.: Grüner bzw. dunkelblauer Annahmenverlauf der jährlichen Zuzüge (oben) bzw. Fortzüge (unten) nach dem Alter und der Zeit von 1994 bis 2050 für die Männer

In der 12. koordinierten Bevölkerungsprognose vom Statistischen Bundesamt [266] werden zwei Annahmenverläufe zum Wanderungssaldo berücksichtigt. Dabei wird eine konstante Anzahl von Fortzügen um eine anzunehmende Anzahl von Zuzügen ergänzt, sodass sich ein Wanderungssaldo von 100.000 bzw. 200.000 Personen ergibt. Die Alters- und Geschlechtsstruktur der Zu- und Fortzüge bleibt dabei unverändert und entspricht einem langjährigen Durchschnitt. Zwei Annahmenverläufe können das zukünftige Entwicklungsspektrum der Migration kaum abdecken, da Zu- und Abwanderung aus dem komplexen, nahezu uneinsehba-



ren Zusammenwirken sozialer, ökonomischer und politisch(-rechtlich)er Bedingungen in den entsprechenden Ziel- und Herkunftsländern resultiert. Folglich sollten in einer Bevölkerungsprognose insbesondere für die Migration mehrere potentielle Entwicklungen flexibel gewichtet werden können, sodass der Unsicherheitsbereich adäquat abgedeckt werden kann. In der Beispielprojektion ergeben sich so schon allein aus den zwei Annahmenverläufen für die direkten Zu- und Fortzüge der Frauen und den drei bzw. zwei Annahmenverläufen für die direkten Zu- bzw. Fortzüge der Männer 24 verschiedene Wanderungssalden (vgl. Abbildung 12.15), die zudem alle flexibel gewichtet werden.

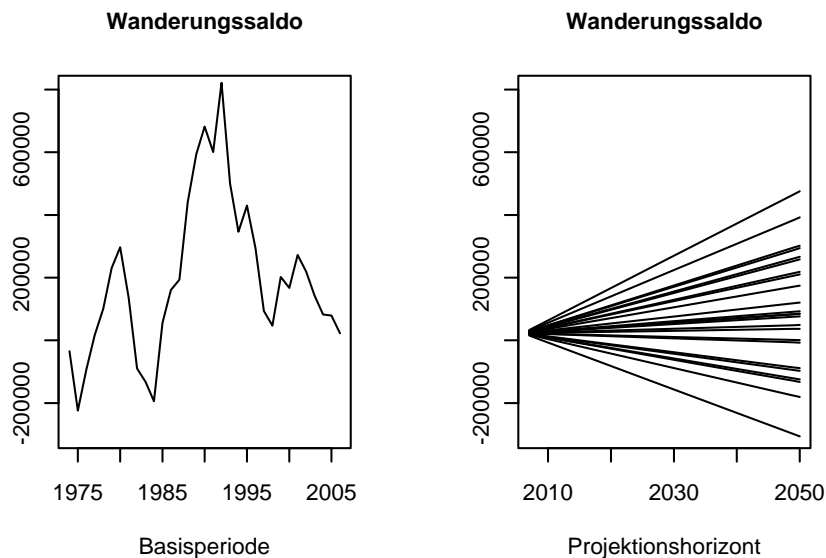


Abbildung 12.15.: Wanderungssalden, die sich aus den Annahmenverläufen  $j$  der jährlichen Zu- und Fortzüge ergeben

#### 12.4.4. Sonstige Annahmen

Nachfolgend sollen die Annahmen für die restlichen Modellparameter der Beispielprojektion kurz beschrieben werden.

Der weibliche bzw. männliche Geschlechteranteil bei Geburt,  $\pi(P_0^F)$  und  $\pi(P_0^M)$ , soll für alle weiblichen und männlichen Subpopulationen konstant dem stabilen Durchschnittswert der vergangenen Jahre von 0,486 bzw. 0,514 entsprechen.

Weiterhin soll für den Modellparameter *Anteil der Säuglingssterbefälle im ersten Halbjahr*,

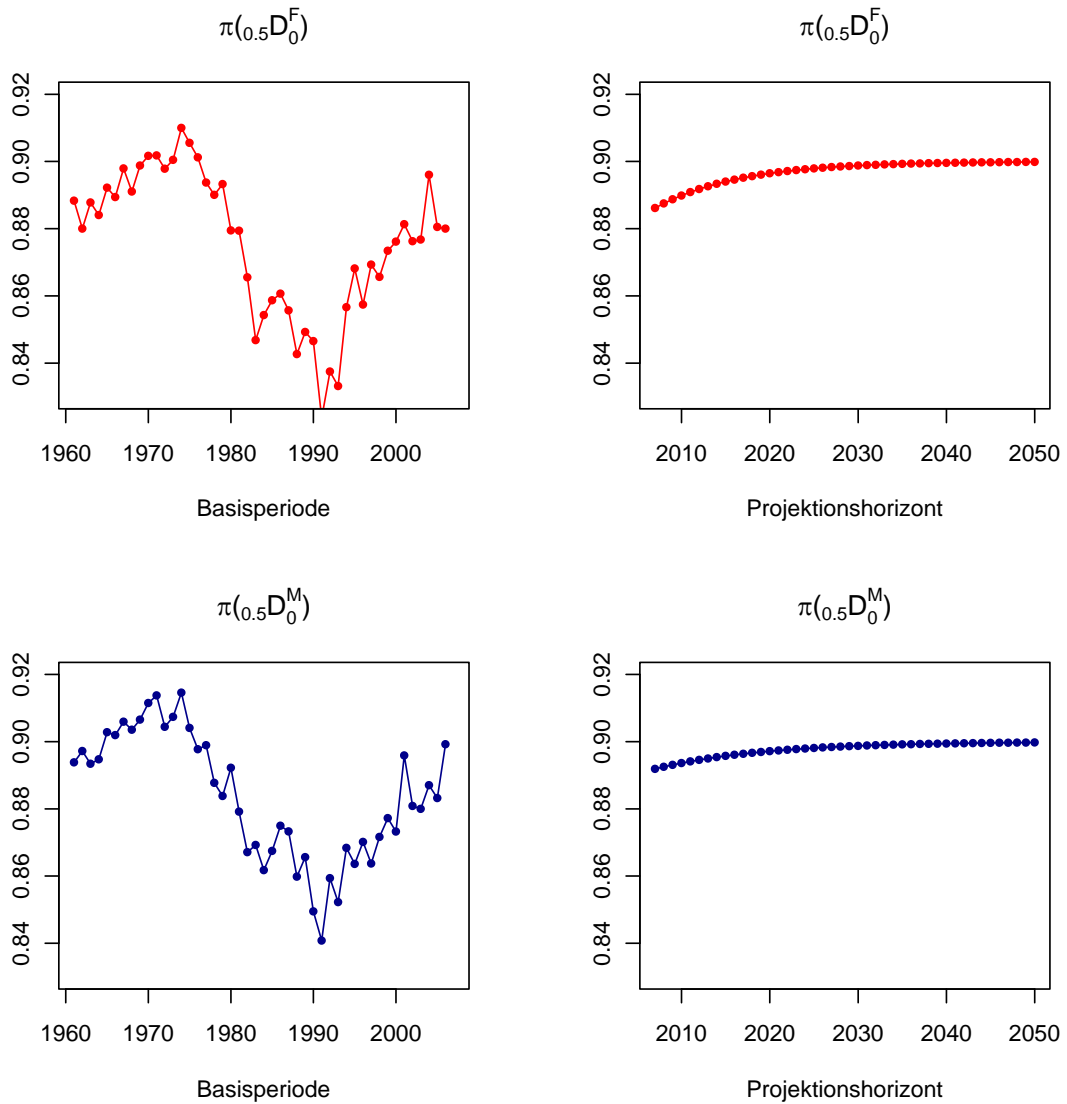


Abbildung 12.16.: Annahmenverlauf zum Anteil der Säuglingssterbefälle im ersten Halbjahr,  $\pi(0.5D_0)$ , für die Mädchen (oben) und die Jungen (unten)

$\pi(0.5D_0)$ , sowohl für die Mädchen als auch für die Jungen (aller Subpopulationen) ein (Wieder)Anstieg auf den Wert von 0,9 angenommen werden (vgl. Abbildung 12.16).<sup>5</sup>

Abschließend ergibt sich die Überlebenswahrscheinlichkeit der 95+-Jährigen für die einzelnen Subpopulationen aus der Differenz von 1 und den entsprechenden prognostizierten alters-

<sup>5</sup>Zur Generierung der konkreten Annahmenwerte ist jeweils ein logistisches Modell geschätzt worden. Die Modellgleichung lautet für die Mädchen:  $\frac{0.9}{1+e^{-2.99246-0.10591t}}$  und für die Jungen:  $\frac{0.9}{1+e^{-2.5095-0.0811t}}$ .

spezifischen Sterbewahrscheinlichkeiten. Folglich stimmt die Anzahl der Annahmenverläufe für die Modellparameter *Überlebende im Alter  $x$*  und *Überlebenswahrscheinlichkeit der obersten, nach oben offenen Altersgruppe*, hier der  $p_{95+}$ , überein.

## 12.5. Berechnung und Analyse der Prognoseergebnisse

Nachdem die Annahmenverläufe für jeden Modellparameter extern generiert worden sind, müssen sie nun in das P3J zur Berechnung einer probabilistischen Bevölkerungsprognose eingegeben werden.

Um die Funktionalität bzw. Vorzüge des PPPM bzw. P3J (gegenüber herkömmlichen Bevölkerungsprognose(modelle)n) zu demonstrieren, sollen sowohl der *Open Type* als auch der *Limited Type* (vgl. Abschnitt 10.2.2 ab Seite 202) mit den Annahmen der Beispielprojektion berechnet werden. Folglich stimmen die ersten beiden Varianten der Beispielprojektion zwar in den einfachen Annahmenverläufen (aus dem Abschnitt 12.4) für jeden Modellparameter überein, allerdings unterscheiden sie sich in der Kombination der Annahmenverläufe: Während im *Open Type* ausnahmslos alle Annahmenverläufe uneingeschränkt miteinander kombiniert werden können, gibt es im *Limited Type* durch die Einführung von Settypen und Sets insofern eine Beschränkung, als dass ausschließlich (Erwartungs- bzw. Theorie-)konsistente Annahmenverläufe plausibel miteinander kombiniert werden können. Darüber hinaus sollen noch zwei weitere Varianten der Beispielprojektion berechnet werden, um auch die Vorgehensweise der repräsentativen Annahmenverläufe (vgl. Abschnitt 10.2.1 ab Seite 200) zu zeigen: In der dritten Variante gibt es (wie häufig auch bei herkömmlichen probabilistischen Bevölkerungsprognosemodellen (vgl. Abschnitt 4.2 ab Seite 57)) lediglich einen repräsentativen Annahmenverlauf anstelle aller einfachen Annahmenverläufe pro Modellparameter, wohingegen in der vierten Variante jeder einfache in einen repräsentativen Annahmenverlauf umgewandelt wird. Die Abbildung 12.17 zeigt diese vier Varianten nochmal im Überblick.

Aufgrund der sehr hohen Anzahl an potentiellen Annahmenverlaufskombinationen in allen vier Varianten der Beispielprojektion — bei der Variante I gibt es insgesamt  $1,132 \cdot 10^{34}$  verschiedene Annahmenverlaufskombinationen — kann von den beiden alternativen Berechnungsmodi (vgl. Abschnitt 10.3.2 ab der Seite 207) nur die Monte-Carlo Simulation tatsächlich ausgeführt werden.

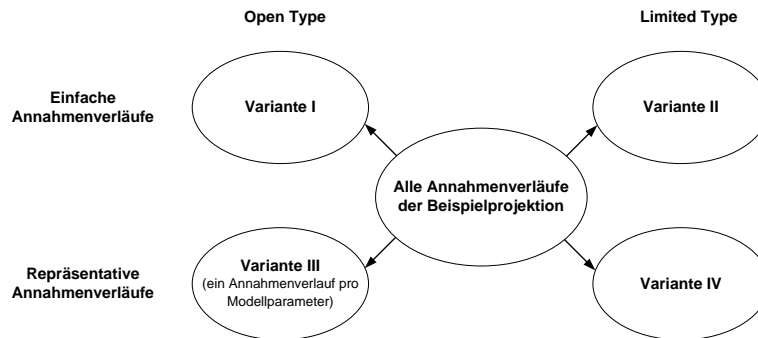


Abbildung 12.17.: Merkmale der vier Varianten der Beispielprojektion

### 12.5.1. Eingabe der Annahmenverläufe ins P3J

Bevor das P3J gestartet werden kann, muss es auf einem Rechner installiert werden. Eine kurze Beschreibung zum Vorgang der Installation des P3J (von der in dieser Arbeit enthaltenen CD-ROM) ist im Anhang A zu finden.

Nach dem Start des P3J (mit einem Doppelklick auf die Datei P3J.bat bzw. P3J.sh im Verzeichnis der in dieser Arbeit enthaltenen CD-ROM) kann eine neue Projektion unter dem Menüpunkt File/New projection angelegt werden (vgl. Abbildung 12.18).

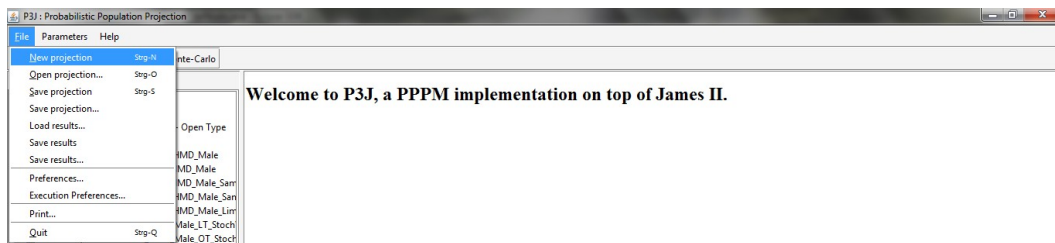


Abbildung 12.18.: Screenshot: Erstellung einer neuen Projektion über den Menüpunkt *New Projection* im P3J.

Im Dialogfenster Define New Projection werden dann der Name der Projektion, die Anzahl der Einzelalter und Nachkommengenerationen (der direkten Zu- und Fortzüge) wie auch die Länge des Projektionshorizonts in Jahren als erste Grundeinstellungen festgelegt (vgl. Abbildung 12.19).

Mit dem Klick auf Add Projection wird die Beispielprojektion im P3J angelegt. Ein Wechsel in das Registerblatt Projection zeigt auf der linken Seite alle Modellparameter in

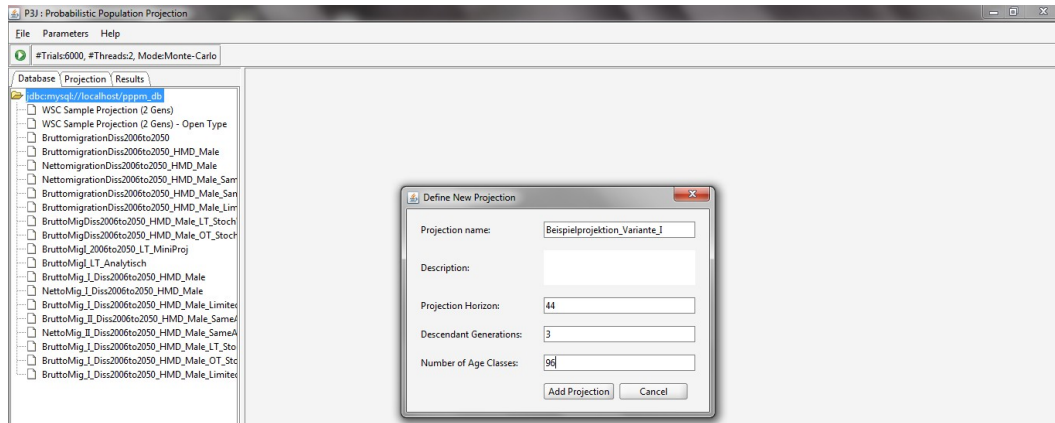


Abbildung 12.19.: Screenshot: Grundeinstellungen für eine neue Projektion im P3J: Im Dialogfenster *Define New Projection* werden der Name der Projektion, die Länge des Projektionshorizonts in Jahren, die Anzahl der Nachkommen- generationen (der direkten Zu- und Fortzüge) wie auch der zu unterscheidenden Einzelalter bestimmt.

einer Baumstruktur. Entsprechend den Grundeinstellungen gibt es jeweils drei Nachkom- mengenerationen für die direkten Zu- und Fortzüge: Descendant Generation 1, 2 und 3 (vgl. Abbildung 12.20).

Um nun die in Abschnitt 12.4 generierten Annahmenverläufe in der Variante 1 der Beispielprojektion einzugeben, wird nachfolgend jeder Modellparameter in der Baum- struktur ausgewählt (bzw. angeklickt) und entsprechend bearbeitet. Um beispielsweise die Annahmenverläufe zur Sterblichkeit der einheimischen Männer einzugeben, wird zunächst der Modellparameter Natives: Mortality (male) bei den einheimischen Männern (im Baum) ausgewählt. Mit einem Klick auf den Button New Assignment auf der rechten Seite des Bildschirms kann dann ein (neuer) Annahmenverlauf eingegeben werden (vgl. Abbildung 12.21).

Für jeden Annahmenverlauf müssen der Name und die Eintrittswahrscheinlichkeit als Ge- neral Information wie auch die Annahmenwerte in dem Registerblatt Data eingegeben werden. Die Dimension der Matrizen für die Annahmenwerte bestimmt sich dabei über die Grund- einstellungen; so bestimmen die Anzahl der Einzelalter und Jahre im Projektionshorizont die Anzahl der Zeilen und Spalten. Die extern generierten Annahmenwerte können u. a. aus

## 12. Beispielhafte Anwendung des PPPM und P3J

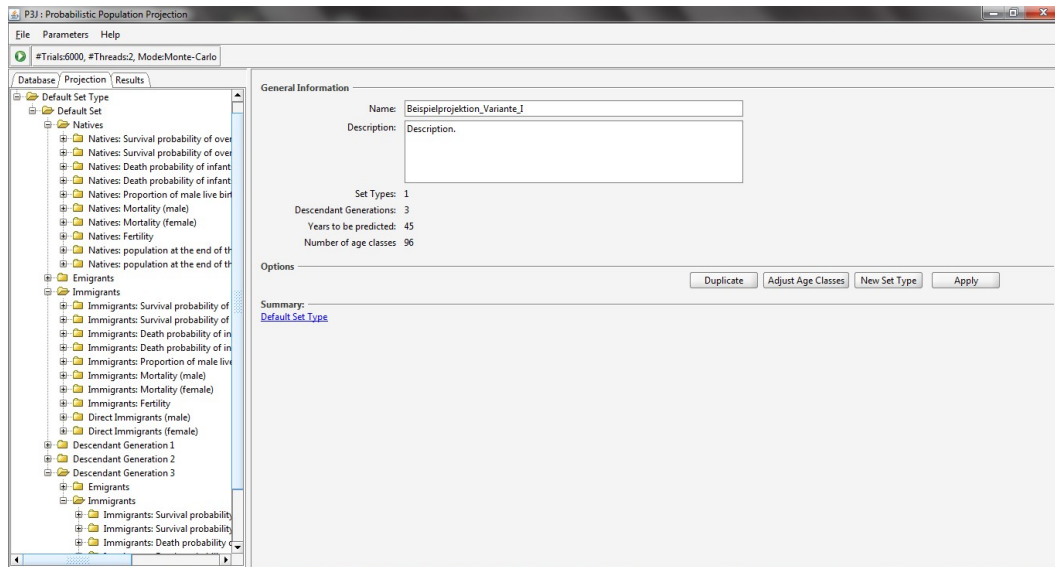


Abbildung 12.20.: Screenshot: Anordnung der Modellparameter und Annahmenverläufe in einer Baumstruktur auf der linken Seite im P3J.

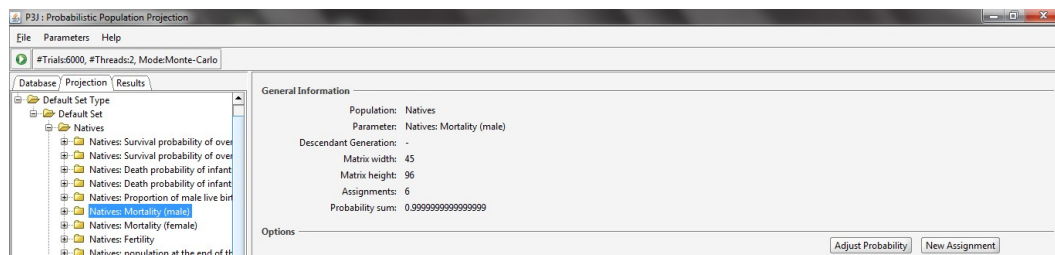


Abbildung 12.21.: Screenshot: Eingabe eines neuen Annahmenverlaufs im P3J über den Button *New Assignment* für einen in der Baumstruktur markierten Modellparameter (in diesem Beispiel für die *Überlebenden im Alter x* der einheimischen Männer).

einem Dateneditor kopiert und im P3J eingefügt werden (vgl. Abbildung 12.22).

**Variante II** Während in der Variante I (bzw. im *Open Type*) alle Modellparameter und Annahmenverläufe in dem Settyp und Set Default zusammengeschlossen sind (vgl. Abbildung 12.20), sollen in der Variante II (bzw. im *Limited Type*) extra Settypen und Sets gebildet werden, um unplausible Annahmenverlaufskombinationen zu eliminieren.

Wenn nun ein Prognostiker erwartet, dass ausschließlich ähnliche bzw. (Theorie-)

## 12. Beispielhafte Anwendung des PPPM und P3J

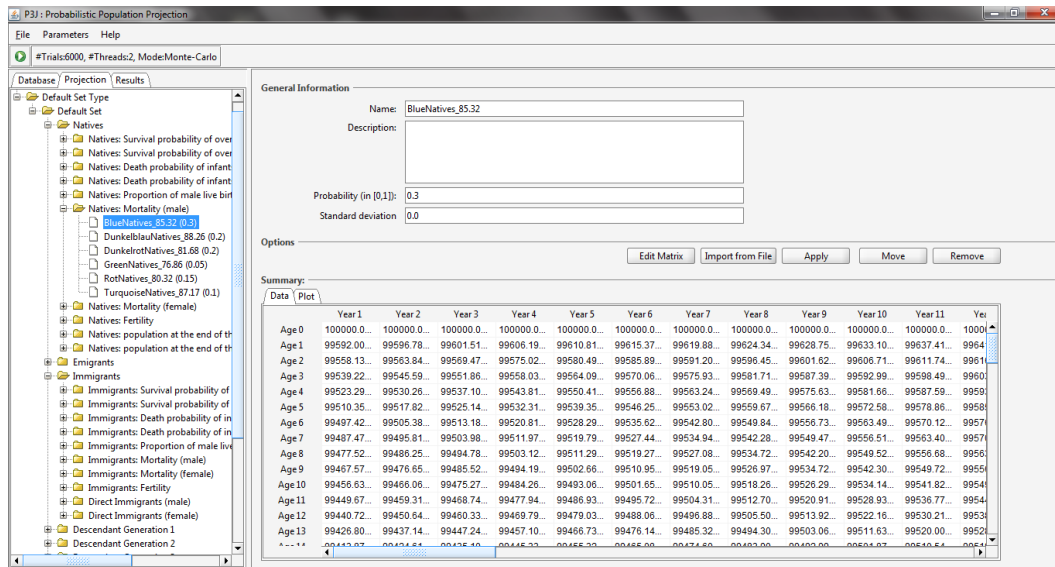


Abbildung 12.22.: Screenshot: Einstellungen für einen neuen Annahmenverlauf im P3J: Dazu werden der Name und die Eintrittswahrscheinlichkeit als *General Information* wie auch die Annahmenwerte in dem Registerblatt *Data* eingegeben.

konsistente Annahmen zur Mortalität, Fertilität und Migration über die entsprechenden Subpopulationen plausibel kombiniert werden können, sollten drei Settypen erstellt werden. Der Settyp Mortalität ( $P_{Mortalität}$ ) schließt dabei die Modellparameter *Überlebende im Alter  $x$*  wie auch die *Überlebenswahrscheinlichkeit der obersten, nach oben offenen Altersgruppe* aller Subpopulationen zusammen, wohingegen der Settyp Fertilität ( $P_{Fertilität}$ ) den Modellparameter *altersspezifische Fertilitätsraten* über alle Subpopulationen in sich vereint. Die Modellparameter *direkte Zuzüge* und *direkte Fortzüge* werden hingegen für Frauen und Männer zusammen im Settyp Migration ( $P_{Migration}$ ) subsumiert.

Im P3J kann mit einem Klick auf den Button New Set Type (vgl. Abbildung 12.20) ein neuer Settyp gebildet werden. Um beispielsweise den Settyp Mortality zu bilden, der sowohl die Überlebenden im Alter  $x$  als auch die Überlebenswahrscheinlichkeit der 95+-Jährigen über alle Subpopulationen vereinen soll, werden im Dialogfenster Create New Set Type die entsprechenden Modellparameter in der rechten Liste markiert und dem neu zu bildenden Settyp mit einem Klick auf den entsprechenden Pfeilbutton (nach links) zugeordnet (vgl. Abbildung 12.23).

## 12. Beispielhafte Anwendung des PPPM und P3J

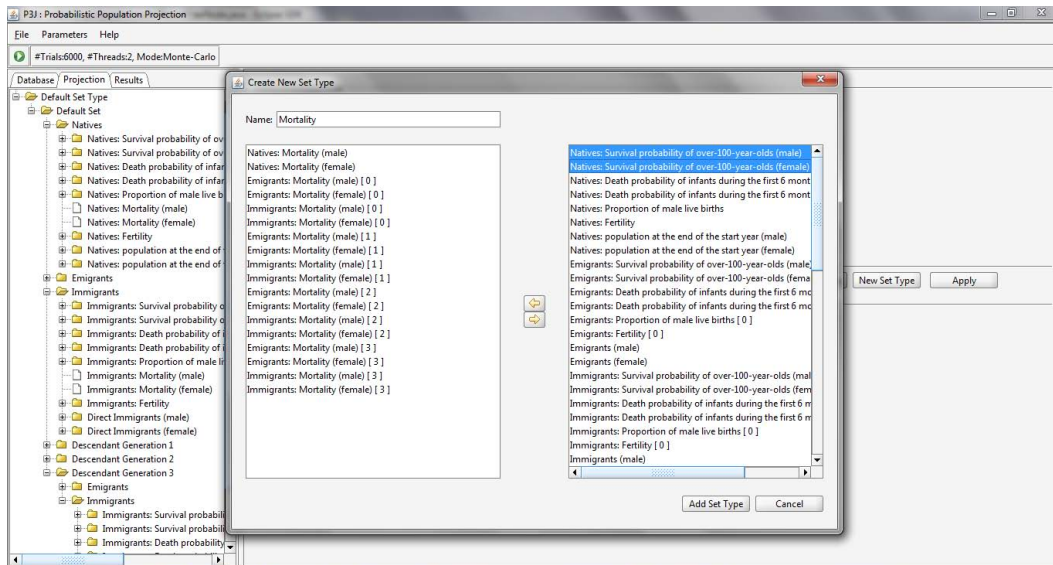


Abbildung 12.23.: Screenshot: Grundeinstellungen für einen neuen Settyp im P3J: Im Dialogfenster *Create New Set Type* werden der Name des Settyps wie auch die in der rechten Liste zu markierenden Modellparameter dem neu zu bildenden Settyp mit einem Klick auf den entsprechenden Pfeilbutton (nach links) zugeordnet.

Für jeden der drei Settypen werden darüber hinaus ein blaues, ein rotes und ein grünes Set ( $\sigma_{blau}^p$ ,  $\sigma_{rot}^p$ ,  $\sigma_{grün}^p$ ) konstruiert, da sich immer die Annahmenverläufe einer Farbfamilie relativ ähnlich sind bzw. einer Erwartung oder Theorie entsprechen. Demzufolge gibt es z.B. für den Settyp Mortalität ein blaues, rotes und grünes Set, die jeweils die blauen, roten und grünen Annahmenverläufe der einzelnen Subpopulationen enthalten. Das blaue bzw. rote Set repräsentiert dabei die ungebrochene bzw. abgeschwächte Fortführung des Langzeittrends (der Mortalitätsreduktion) in die Zukunft. Das grüne Set enthält hingegen Annahmenverläufe, die eine Konstanz der Sterblichkeit aufgrund der Zunahme bestimmter Risikofaktoren (wie z. B. Adipositas) unterstellen.

Im P3J können Sets für einen Settypen mit einem Klick auf den Button New Set (vgl. Abbildung 12.24) angelegt werden, wobei der Name des Sets wie auch dessen Eintrittswahrscheinlichkeit in dem anschließend erscheinenden Dialogfenster Create New Set eingegeben werden müssen (vgl. Abbildung 12.25).



## 12. Beispielhafte Anwendung des PPPM und P3J

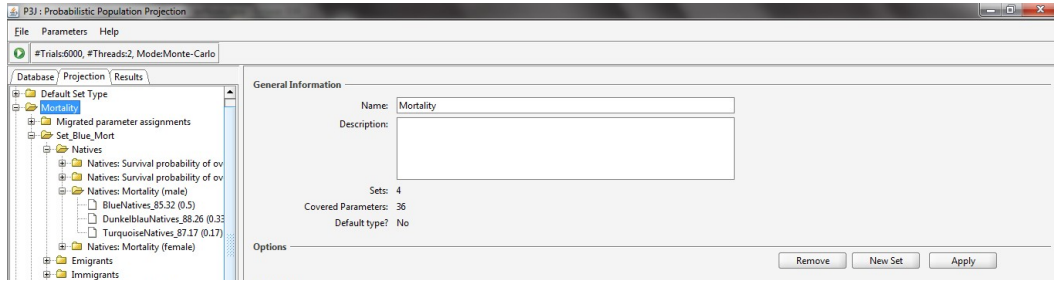


Abbildung 12.24.: Screenshot: Erstellung eines neuen Sets im P3J über den Button *New Set* für einen in der Baumstruktur markierten Settyp (in diesem Beispiel für den Settyp Mortalität).

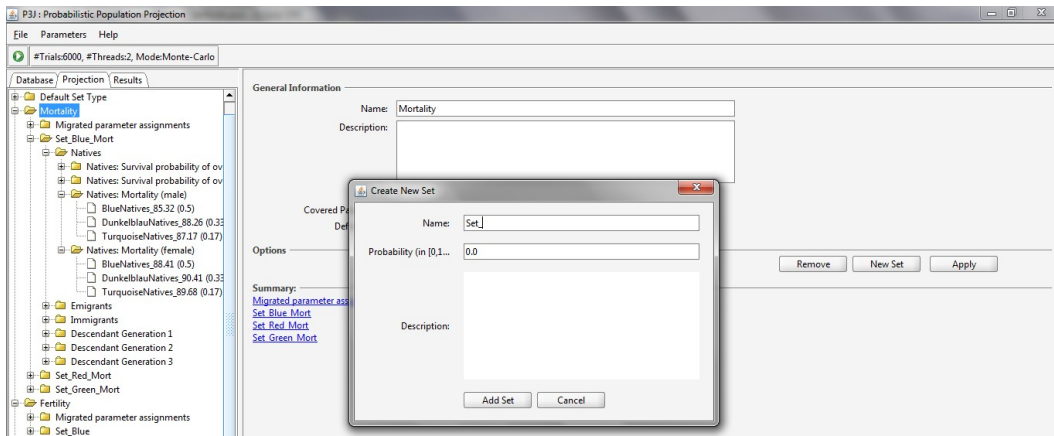


Abbildung 12.25.: Screenshot: Einstellungen für ein neues Set im P3J: Im Dialogfenster *Create New Set* werden der Name, die Eintrittswahrscheinlichkeit wie auch eine Beschreibung des neuen Sets eingegeben.

Die Tabellen 12.7 bis 12.9 geben für jeden der drei Settypen ( $P_{Mortalität}$ ,  $P_{Fertilität}$ ,  $P_{Migration}$ ) einen Überblick über die Sets ( $S_{Mortalität}$ ,  $S_{Fertilität}$ ,  $S_{Migration}$ ) und die jeweils darin zusammenzuschließenden Annahmenverläufe ( $A_{\sigma}^{Mortalität}$ ,  $A_{\sigma}^{Fertilität}$ ,  $A_{\sigma}^{Migration}$ ) mit den entsprechenden Eintrittswahrscheinlichkeiten ( $pr_{\sigma}^p$  und  $pr_{j,\sigma}^p$ ). Die Annahmenverläufe werden dabei für die Settypen Mortalität, Fertilität und Migration durch die Lebenserwartung bei Geburt, die Total Fertility Rate und die direkten Zu- und Fortzüge im Zieljahr repräsentiert. Die Eintrittswahrscheinlichkeiten für die Annahmenverläufe,  $pr_{j,\sigma}^p$ , werden von der Variante I (bzw. vom *Open Type*) in die Variante II (bzw. den *Limited Type*) umgerechnet, sodass sie weitestgehend übereinstimmen.<sup>6</sup>

<sup>6</sup>Eine Ausnahme stellt das rote Set des Settyps Migration dar; hier werden grüne Annahmenverläufe für die

## 12. Beispielhafte Anwendung des PPPM und P3J

	$a_{j,\sigma_{\text{blau}}}^{\text{Fertilität}(TFR)}$	$pr_{j,\sigma_{\text{blau}}}^{\text{Fertilität}}$	$a_{j,\sigma_{\text{rot}}}^{\text{Fertilität}(TFR)}$	$pr_{j,\sigma_{\text{rot}}}^{\text{Fertilität}}$	$a_{j,\sigma_{\text{grün}}}^{\text{Fertilität}(TFR)}$	$pr_{j,\sigma_{\text{grün}}}^{\text{Fertilität}}$
E	1,59*	0,33	1,419*	1	1,309*	1
	1,507*	0,67				
SZ	2	0,4	1,84	1	1,74	1
	1,92	0,6				
SZG(1)	1,9	0,4	1,69	1	1,54	1
	1,81	0,6				
SZG(2)	1,59*	1	1,46	1	1,31*	1
SZG(3)	1,59*	0,33	1,419*	1	1,309*	1
	1,507*	0,67				
SFz	1,59*	0,33	1,419*	1	1,309*	1
	1,507*	0,67				
SFzG(1)	1,59*	0,33	1,419*	1	1,309*	1
	1,507*	0,67				
SFzG(2)	1,59*	0,33	1,419*	1	1,309*	1
	1,507*	0,67				
SFzG(3)	1,59*	0,33	1,419*	1	1,309*	1
	1,507*	0,67				
$pr_{\sigma}^{\text{Fertilität}}$	0,15		0,6		0,25	

Tabelle 12.7.: Settyp Fertilität in der Variante II der Beispielprojektion: Annahmenverläufe  $a_j$  der Sets  $\sigma$  im Settyp Fertilität und ihre Eintrittswahrscheinlichkeiten  $pr_{j,\sigma}$ ; die mit einem \* markierten Annahmenverläufe entsprechen jeweils einem Annahmenverlauf der einheimischen Frauen.

**Variante III und IV** Im PPPM bzw. P3J können nicht nur einfache, sondern auch repräsentative Annahmenverläufe mit stochastischer Variation eingegeben werden. Bei der Eingabe eines repräsentativen Annahmenverlaufs muss dann (im Vergleich zu den einfachen Annahmenverläufen) neben dem Namen, der Eintrittswahrscheinlichkeit und den (repräsentativen) Annahmenwerten auch noch die Standardabweichung zur Bestimmung des Ausmaßes der stochastischen Variation eingegeben werden. Bei den einfachen Annahmenverläufen beträgt diese Standardabweichung null, wohingegen sie bei den repräsentativen Annahmenverläufen größer als null ist; in den Varianten III und IV soll sie (immer) 0,005 betragen.

Im P3J wird die Standardabweichung als General Information im Feld Standard deviation eingetragen (vgl. Abbildung 12.22).

---

Modellparameter eingeschlossen, für die es keine roten Annahmenverläufe gibt, sodass die Eintrittswahrscheinlichkeit dieser grünen Annahmenverläufe in der Variante II gegenüber der Variante I etwas zunimmt.

## 12. Beispielhafte Anwendung des PPPM und P3J

	$a_{j,\sigma_{blau}}^{Mortalität(e_0)}$	$pr_{j,\sigma_{blau}}^{Mortalität}$	$a_{j,\sigma_{rot}}^{Mortalität(e_0)}$	$pr_{j,\sigma_{rot}}^{Mortalität}$	$a_{j,\sigma_{grün}}^{Mortalität(e_0)}$	$pr_{j,\sigma_{grün}}^{Mortalität}$
E Frauen	90,41*	0,33	85,79*	0,57	82,03*	1
	89,68*	0,17	84,76*	0,43		
	88,41*	0,5				
E Männer	88,26*	0,33	81,68*	0,57	76,86*	1
	87,17*	0,17	80,32*	0,43		
	85,32*	0,5				
SZ Frauen	92,63	0,57	88,05	0,6	84,78	1
	90,71	0,43	87,09	0,4		
SZ Männer	91,68	0,57	85,2	0,6	80,88	1
	88,91	0,43	83,91	0,4		
SZG(1) Frauen	91,61	0,36	87,07	0,6	83,47	1
	89,59	0,64	86,16	0,4		
SZG(1) Männer	90,2	0,36	83,37	0,6	78,89	1
	87,22	0,64	82,27	0,4		
SZG(2) Frauen	90,41*	0,36	85,79*	0,6	82,03*	1
	89,68*	0,07	84,76*	0,4		
	88,41*	0,57				
SZG(2) Männer	88,26*	0,36	81,68*	0,6	76,86*	1
	87,17*	0,07	80,32*	0,4		
	85,32*	0,57				
SZG(3) Frauen	90,41*	0,33	85,79*	0,57	82,03*	1
	89,68*	0,17	84,76*	0,43		
	88,41*	0,5				
SZG(3) Männer	88,26*	0,33	81,68*	0,57	76,86*	1
	87,17*	0,17	80,32*	0,43		
	85,32*	0,5				
SFz Frauen	90,41*	0,33	85,79*	0,57	82,03*	1
	89,68*	0,17	84,76*	0,43		
	88,41*	0,5				
SFz Männer	88,26*	0,33	81,68*	0,57	76,86*	1
	87,17*	0,17	80,32*	0,43		
	85,32*	0,5				
SFzG(1) Frauen	90,41*	0,33	85,79*	0,57	82,03*	1
	89,68*	0,17	84,76*	0,43		
	88,41*	0,5				
SFzG(1) Männer	88,26*	0,33	81,68*	0,57	76,86*	1
	87,17*	0,17	80,32*	0,43		
	85,32*	0,5				
SFzG(2) Frauen	90,41*	0,33	85,79*	0,57	82,03*	1
	89,68*	0,17	84,76*	0,43		
	88,41*	0,5				
SFzG(2) Männer	88,26*	0,33	81,68*	0,57	76,86*	1
	87,17*	0,17	80,32*	0,43		
	85,32*	0,5				
SFzG(3) Frauen	90,41*	0,33	85,79*	0,57	82,03*	1
	89,68*	0,17	84,76*	0,43		
	88,41*	0,5				
SFzG(3) Männer	88,26*	0,33	81,68*	0,57	76,86*	1
	87,17*	0,17	80,32*	0,43		
	85,32*	0,5				
$pr_{\sigma}^{Mortalität}$	0,675		0,275		0,05	

Tabelle 12.8.: Settyp Mortalität in der Variante II der Beispielpjektion: Annahmenverläufe  $a_j$  der Sets  $\sigma$  im Settyp Mortalität und ihre Eintrittswahrscheinlichkeiten  $pr_{j,\sigma}$ ; die mit einem \* markierten Annahmenverläufe entsprechen jeweils einem Annahmenverlauf der Einheimischen.

## 12. Beispielhafte Anwendung des PPPM und P3J

	$a_{j,\sigma_{blau}}^{Migration}$	$pr_{j,\sigma_{blau}}^{Migration}$	$a_{j,\sigma_{rot}}^{Migration}$	$pr_{j,\sigma_{rot}}^{Migration}$	$a_{j,\sigma_{grün}}^{Migration}$	$pr_{j,\sigma_{grün}}^{Migration}$
Z Frauen	442.380	1	224.751	1	224.751	1
Z Männer	560.870	1	477.226	1	351.783	1
Fz Frauen	375.238	1	201.577	1	201.577	1
Fz Männer	507.567	1	325.975	1	325.975	1
$pr_{\sigma}^{Migration}$	0,15		0,1		0,75	

Tabelle 12.9.: Settyp Migration in der Variante II der Beispielprojektion: Annahmenverläufe  $a_j$  der Sets  $\sigma$  im Settyp Migration und ihre Eintrittswahrscheinlichkeiten  $pr_{j,\sigma}$ .

Subpopulation	Fertilität	Mortalität	Migration
E	rot	blau	-
SZ	rot	blau	grün
SZG(1)	rot	blau	-
SZG(2)	rot	blau	-
SZG(3)	rot	blau	-
SFz	rot	blau	grün
SFzG(1)	rot	blau	-
SFzG(2)	rot	blau	-
SFzG(3)	rot	blau	-

Tabelle 12.10.: Repräsentative Annahmenverläufe in der Variante III der Beispielprojektion

Die Tabelle 12.10 gibt einen Überblick über die ausgewählten repräsentativen Annahmenverläufe in der Variante III für jeden Modellparameter; die repräsentativen Annahmenverläufe sind dabei so ausgewählt worden, dass sie die potentiellen Entwicklungen so gut wie möglich einfangen; folglich stellen bei dieser Auswahl eine mittlere Lage des Annahmenverlaufs in der jeweiligen Annahmenverteilung wie auch die Höhe der jeweiligen Eintrittswahrscheinlichkeit die ausschlaggebenden Kriterien dar.

Die Abbildungen 12.26 und 12.27 zeigen am Beispiel der Mortalität der einheimischen Männer, wie sich die Annahmenverteilung in den Varianten III und IV im Zieljahr aufgrund der Umwandlung der einfachen in repräsentative Annahmenverläufe verändert.

In der Variante III soll der blaue, nun repräsentative Annahmenverlauf (zusammen mit seinen stochastischen Abweichungen<sup>7)</sup> alle einfachen Annahmenverläufe der Variante I

<sup>7</sup>Im Anhang C ist ein Beispiel für eine stochastische Variation des blauen repräsentativen Annahmenverlaufs

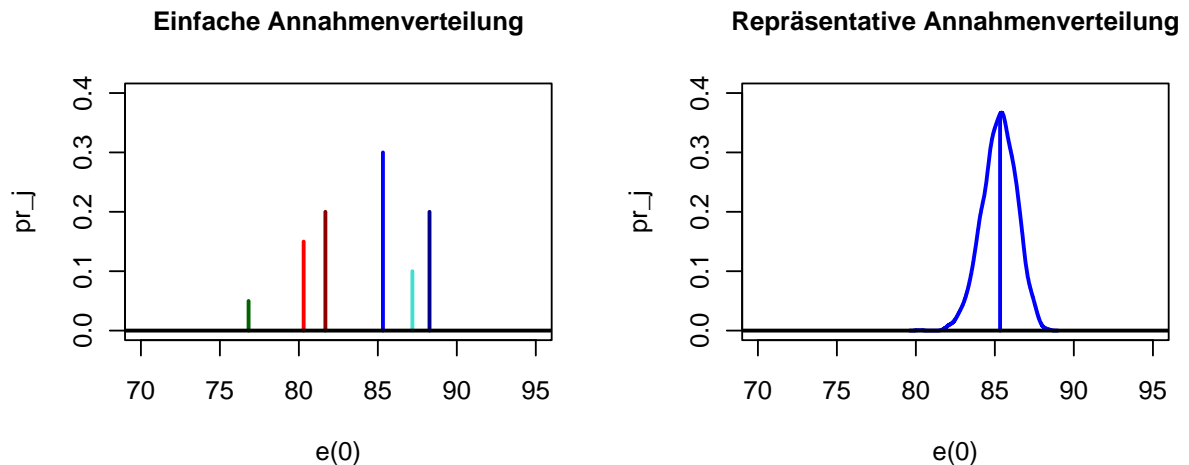


Abbildung 12.26.: *Einfache* und *repräsentative* Annahmenverteilung zur Mortalität der einheimischen Männer in den Varianten I (links) und III (rechts) der Beispielprojektion im Zieljahr

ersetzen. Ein Vergleich dieser beiden Annahmenverteilungen zeigt jedoch, dass die Annahmenverteilung der Variante III nur einen Bruchteil der Annahmenverteilung der Variante I abdecken kann. Folglich kann die heterogene Annahmenverteilung der Variante I, die sich aus mehreren einfachen, flexibel gewichteten Annahmenverläufen zusammensetzt, nicht (wie häufig in herkömmlichen probabilistischen Bevölkerungsprognosemodellen) durch einen repräsentativen Annahmenverlauf mit stochastischer Variation ersetzt werden.

Aus diesem Grund werden in der Variante IV alle einfachen Annahmenverläufe in repräsentative Annahmenverläufe mit stochastischer Variation umgewandelt. Mit der dadurch entstehenden Annahmenverteilung (in der Variante IV) kann die ursprüngliche Annahmenverteilung der Variante I nicht nur abgedeckt, sondern auch noch leicht erweitert werden. So hebt die stochastische Variation der repräsentativen Annahmenverläufe die klaren Grenzen zwischen den verschiedenen einfachen Annahmenverläufen auf und deckt somit die Annahmenverteilung noch dichter (und breiter) ab.

### 12.5.2. Berechnung der Prognoseergebnisse im P3J

Für jede Variante der Beispielprojektion werden 6.000 Prognosedurchläufe mit der Monte-Carlo Simulation durchgeführt.

---

zur Sterblichkeit der einheimischen Männer abgebildet.

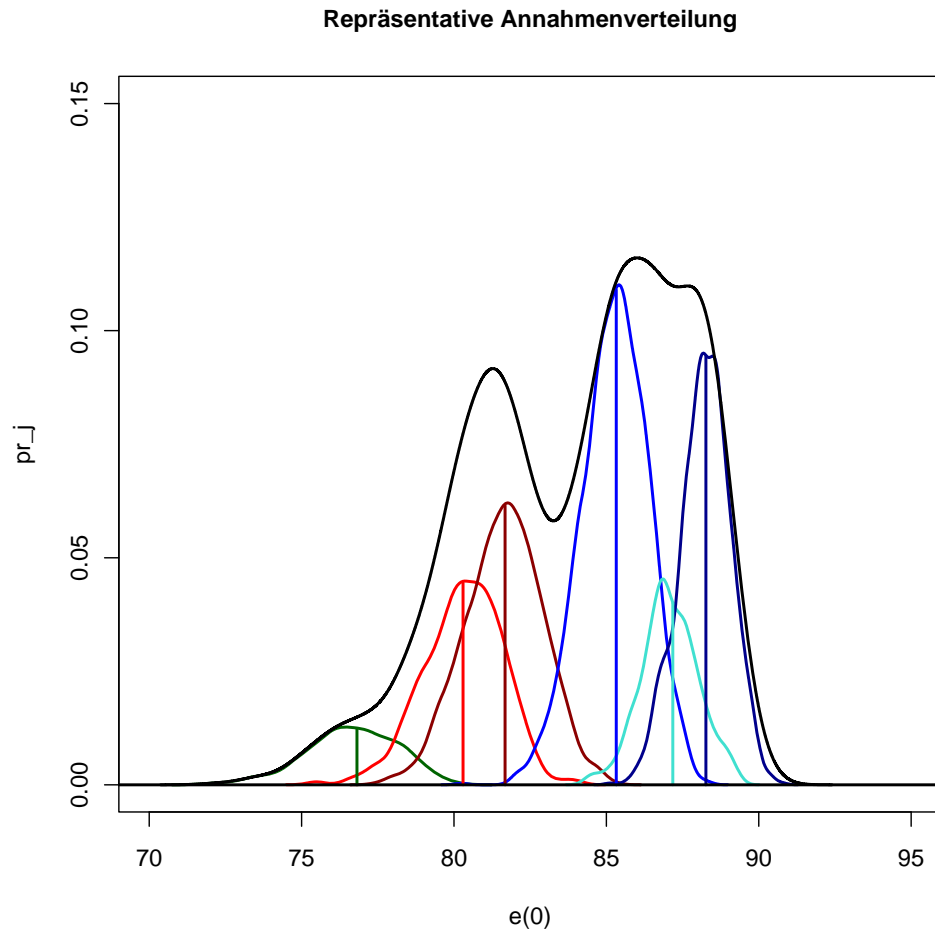


Abbildung 12.27.: Komplexe *repräsentative* Annahmenverteilung zur Mortalität der einheimischen Männer in der Variante IV der Beispielprojektion im Zieljahr; die gesamte Annahmenverteilung (schwarz) setzt sich aus den einzelnen Annahmenverteilungen der sechs repräsentativen Annahmenverläufe zusammen; die vertikalen Linien stellen die eingegebenen (mittleren) Annahmenwerte dar.

Die Anzahl der Prognosedurchläufe und der Berechnungsmodus werden im P3J in dem Dialogfenster Edit Execution Preferences bestimmt, das über den Menüpunkt File/Preferences... aufgerufen wird (vgl. Abbildungen 12.28 und 12.29). Eine Simulation wird dann mit einem Klick auf den Button Run (weißer Pfeil nach rechts in grünem Kreis) gestartet.

Nach einer Simulation können die Prognoseergebnisse wesentlicher Ergebnisgrößen als Roh-

## 12. Beispielhafte Anwendung des PPPM und P3J

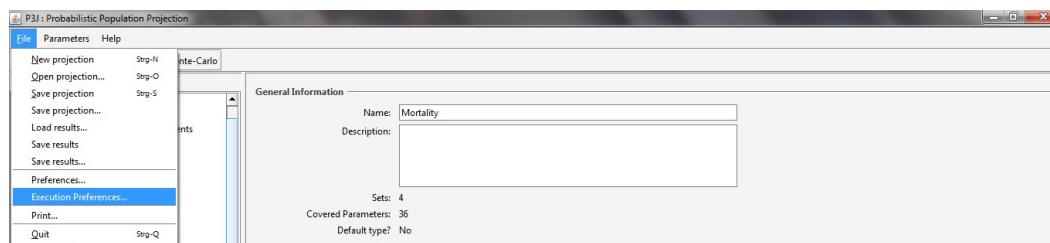


Abbildung 12.28.: Screenshot: Bestimmung der Anzahl der Prognosedurchläufe und des Berechnungsmodus über den Menüpunkt *Execution Preferences* im P3J.

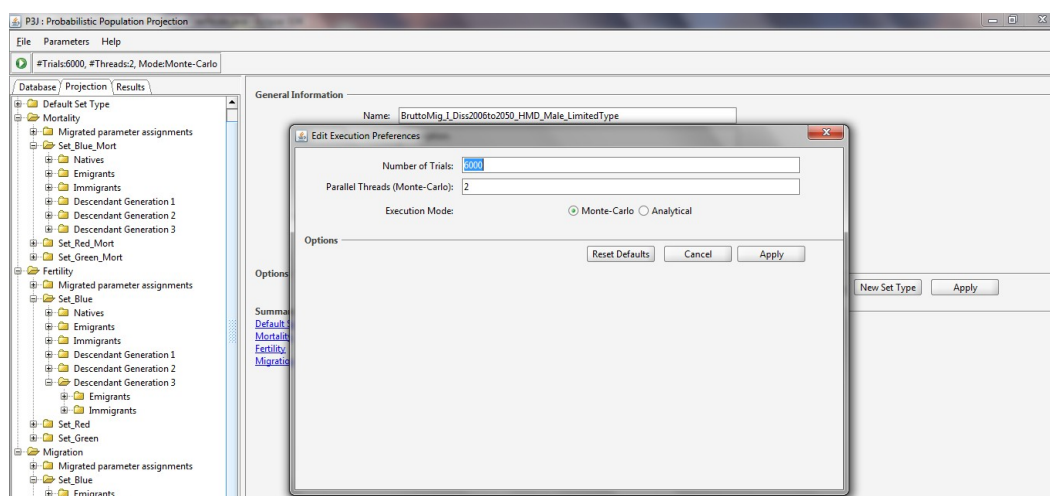


Abbildung 12.29.: Screenshot: Einstellung der Anzahl der Prognosedurchläufe und des Berechnungsmodus im Dialogfenster *Edit Execution Preferences* im P3J.

daten exportiert oder/und zur Erstellung eines Ergebnisreports aggregiert und in einem selbst zu bestimmenden Zielverzeichnis gespeichert werden. Dazu wird im Registerblatt Results im Baum auf Overall Results und anschließend auf den Button *Export Data* bzw. *Generate Report* geklickt (vgl. Abbildung 12.30).

**Vergleich der Varianten I bis IV** Bei der Auswertung der Rohdaten kann beispielsweise die prognostizierte Entwicklung der Gesamtbevölkerung über den gesamten Projektionshorizont (vgl. Abbildung 12.31) oder auch als kumulierte (aufsteigend sortierte) Wahrscheinlichkeitsverteilung (vgl. Abbildung 12.32) bzw. als entsprechende Dichteverteilung (vgl. Abbildung 12.33) im Zieljahr 2050 für alle vier Varianten der Beispielprojektion verglichen werden.

## 12. Beispielhafte Anwendung des PPPM und P3J

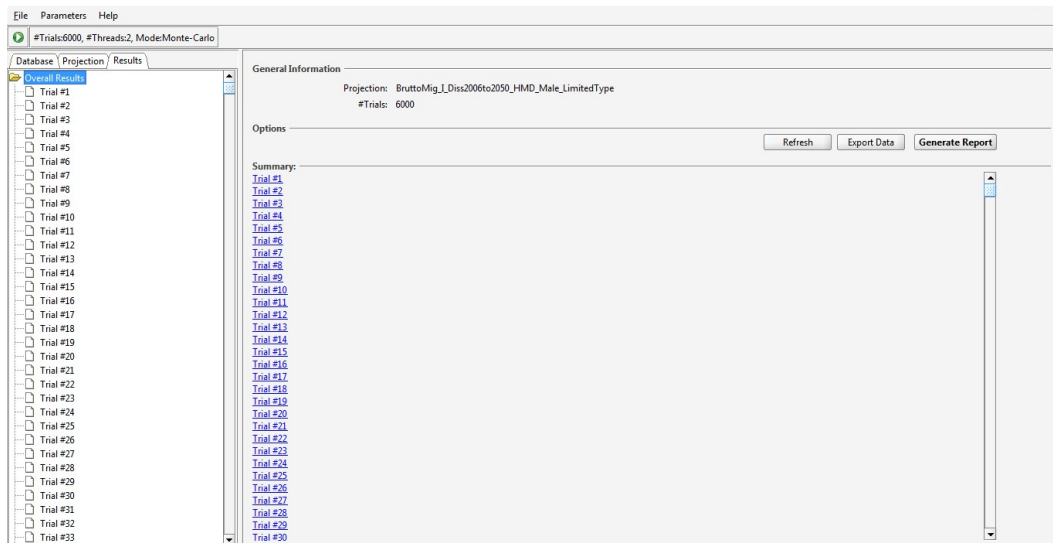


Abbildung 12.30.: Screenshot: Exportieren bzw. aggregieren der Prognoseergebnisse über den Button *Export Data* bzw. *Generate Report* im P3J, wobei in der Baumstruktur der Ordner *Overall Results* markiert werden muss.

Die Abbildung 12.31 zeigt generell, dass die Gesamtbevölkerung in allen vier Varianten abnimmt und die Prognoseunsicherheit mit zunehmender Länge des Projektionshorizonts wächst. Allerdings unterscheiden sich die vier Varianten deutlich in dem Ausmaß der Streuung, wohingegen sie im Median beinahe übereinstimmen.

Der Vergleich der Varianten I und II zeigt, dass die Ergebnisverteilung der Gesamtbevölkerung für den *Limited Type* deutlich schmaler ist als für den *Open Type*. Demzufolge bewirkt die Elimination unplausibler Kombinationen im *Limited Type* eine deutliche Reduktion des Unsicherheitsbereichs im Vergleich zum *Open Type*. So umfasst die Gesamtbevölkerung im Zieljahr mit einer Wahrscheinlichkeit von 95 Prozent zwischen 61.661.136 und 69.205.329 Personen in der Variante II und zwischen 60.201.707 und 78.926.850 Personen in der Variante I; folglich ist der Unsicherheitsbereich in der Variante I durch die unplausiblen Kombinationen um etwa 52 Prozent größer als in der Variante II.

Darüber hinaus bewirkt die Elimination der unplausiblen Kombinationen in der Variante II eine Verringerung der potentiellen Kombinationen,  $N_A$ , gegenüber der Variante I. Sind in der Variante I  $1,132 \cdot 10^{34}$  verschiedene Annahmenkombinationen möglich, sind es in der Variante II nur noch  $1,499 \cdot 10^{18}$  Kombinationen. Dies schlägt sich in der Form der kumulier-



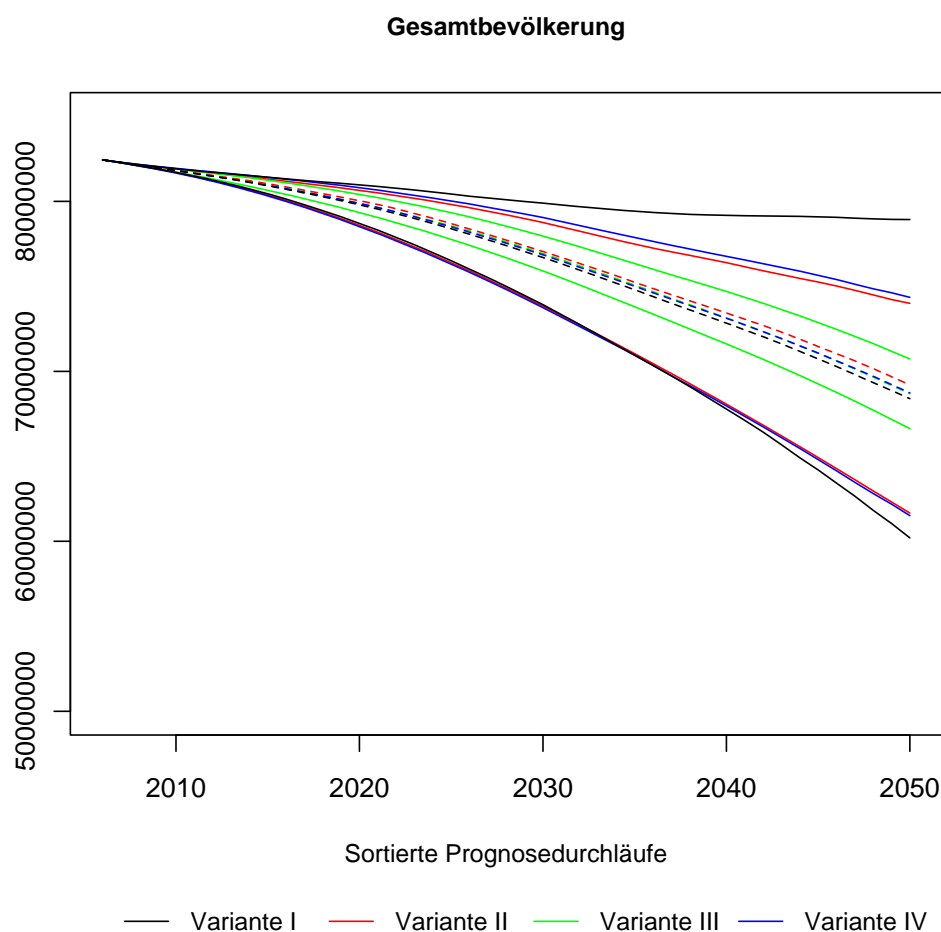


Abbildung 12.31.: Entwicklung der Gesamtbevölkerung über den Projektionshorizont für die Quantile 0,025, 0,5 (gestrichelte Linie) und 0,975 in den Varianten I bis IV der Beispielprojektion

ten Wahrscheinlichkeitsverteilung im Zieljahr (in der Abbildung 12.32) nieder: Während die Verteilung in der Variante I aufgrund der Vielzahl der potentiellen Kombinationen glatt ist, ist sie in der Variante II aufgrund der deutlich geringeren Anzahl potentieller Kombinationen stufenförmig ausgeprägt.

Der Vergleich aller vier Varianten zeigt zudem, dass nur *mehrere* repräsentative Annahmenverläufe mit stochastischer Variation pro Modellparameter die heterogene Ergebnisverteilung der Variante II (mit einer Vielzahl ausschließlich plausibler Zukunftsentwicklungen) annähernd wiedergeben können. So stimmen die Ergebnisverteilungen für die Gesamtbevölkerung in der Variante II und IV im Niveau wie auch im Verlauf nahezu überein,

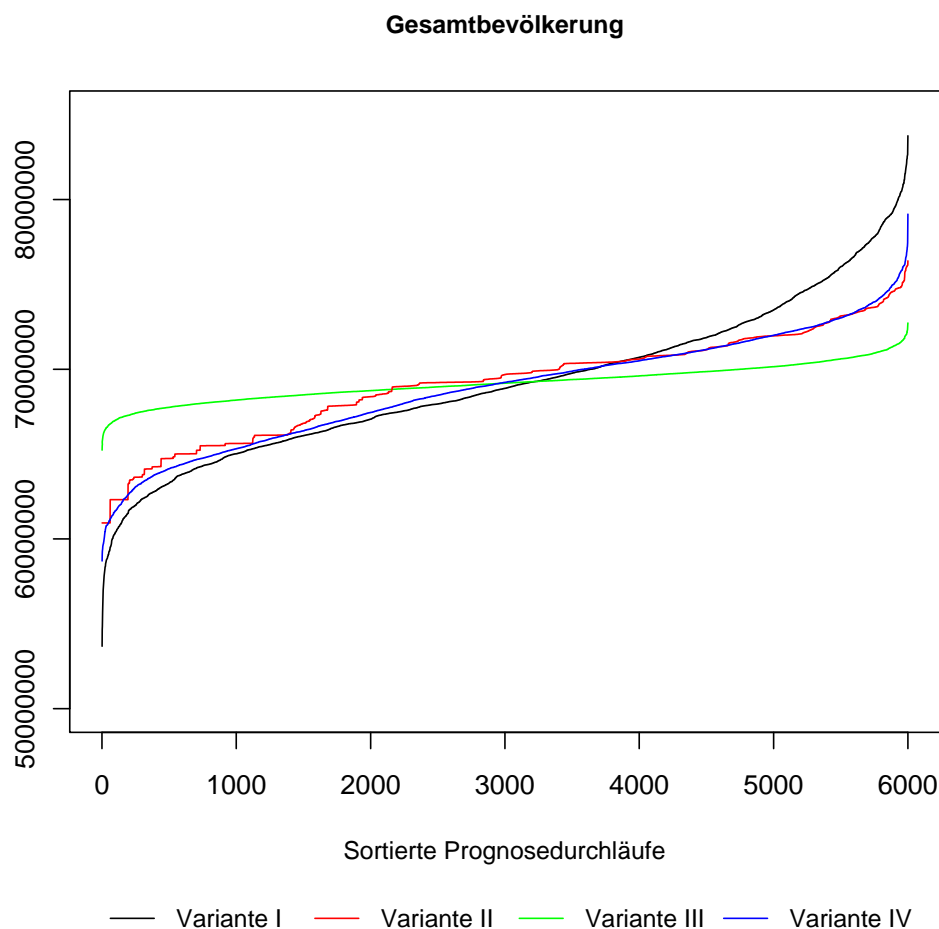


Abbildung 12.32.: Kumulierte Wahrscheinlichkeitsverteilung der Gesamtbevölkerung im Zieljahr in den Varianten I bis IV der Beispielprojektion

allerdings ist die kumulierte Wahrscheinlichkeitsverteilung im Zieljahr in der Variante IV sowohl glatter als auch etwas breiter gestreut als in der Variante II. Die stochastische Variation der repräsentativen Annahmenverläufe bewirkt zwar in der Variante IV eine dichtere Ergebnisverteilung, sodass die stufenförmige (kumulierte) Wahrscheinlichkeitsverteilung der Variante II ausgefüllt und erweitert wird, jedoch (möglicherweise auch) zuungunsten der Plausibilität der Prognoseergebnisse. So können z. B. nicht alle zusätzlich in der Variante IV zu berücksichtigenden Zukunftsszenarien den Erwartungen eines Prognostikers entsprechen, sodass diese zusätzlichen unplausiblen Ergebnisverläufe die eigentliche Ergebnisverteilung verzerren und die Prognosegenauigkeit schmälern können. Dieser Effekt kann mit einem steigenden Wert der Standardabweichung (für die stochastische Variation) zunehmen.

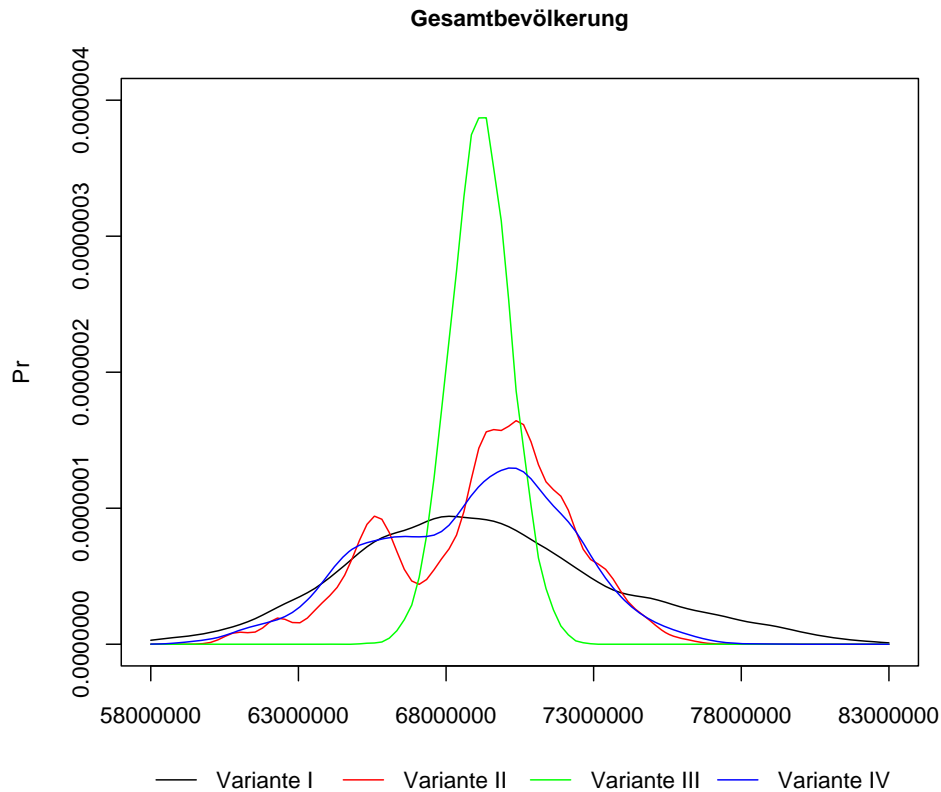


Abbildung 12.33.: Dichteverteilung der Gesamtbevölkerung im Zieljahr in den Varianten I bis IV der Beispielprojektion

Daneben ist zu erkennen, dass die Wahrscheinlichkeitsverteilung der Variante III deutlich von denen der anderen drei Varianten abweicht. Der Ersatz aller mit immer nur *einem* repräsentativen Annahmenverlauf pro Modellparameter führt hier zu einem Ausschluss zahlreicher potentieller Zukunftsszenarien, sodass die resultierende Wahrscheinlichkeitsverteilung fälschlicherweise deutlich schmaler (bzw. die Varianz deutlich kleiner) ist als in den anderen Varianten. Folglich kann die eigentlich heterogene Annahmen- und Ergebnisverteilung der Variante II nur unvollständig mit einem repräsentativen Annahmenverlauf pro Modellparameter in der Variante III (bzw. mit entsprechenden herkömmlichen probabilistischen Bevölkerungsprognoseansätzen, die verschiedene Annahmenverläufe aus zufällig normalverteilten Abweichungen um einen vorgegebenen mittleren Annahmenverlauf generieren) wiedergegeben werden.

**Vergleich der Variante II mit der 12. koordinierten Bevölkerungsprognose** Der Vergleich der Prognoseergebnisse zwischen der Variante II der Beispielprojektion und denen

der 12. koordinierten Bevölkerungsprognose vom Statistischen Bundesamt (vgl. Abbildung 12.34) offenbart ebenfalls einige Vorzüge des PPPM bzw. P3J.

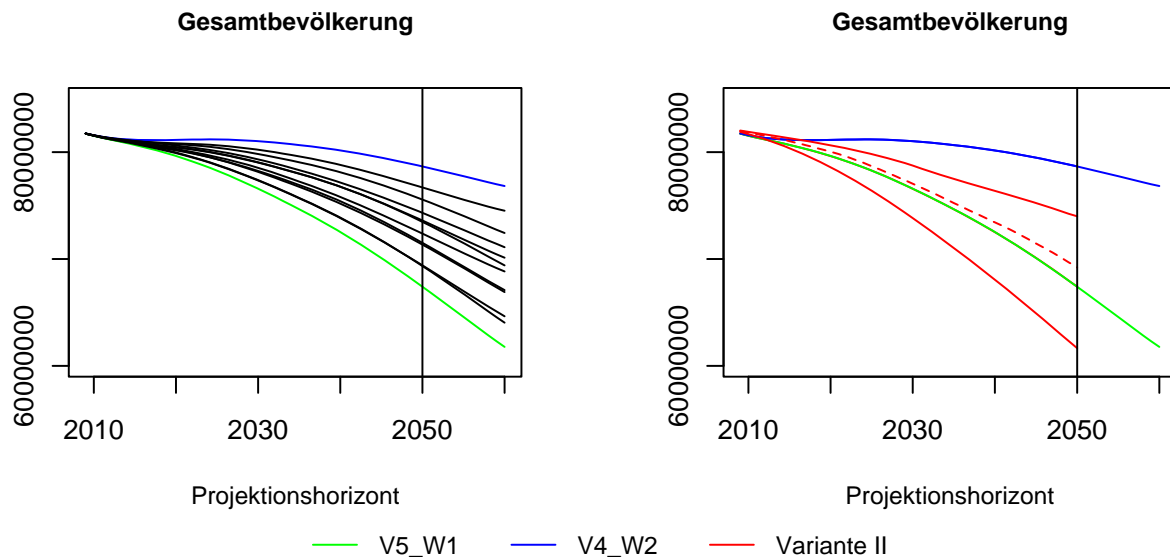


Abbildung 12.34.: Entwicklung der Gesamtbevölkerung von 2009 bis 2060 in den 12 Varianten der 12. koordinierten Bevölkerungsprognose vom Statistischen Bundesamt (links) im Vergleich zur Variante II der Beispielprojektion (rechts); die grüne und blaue Linie kennzeichnen die Varianten V5\_W1 und V4\_W2 von der 12. koordinierten Bevölkerungsprognose und die roten Linien entsprechen den Quantilen 0,025, 0,5 (gestrichelte Linie) und 0,975 der Variante II der Beispielprojektion; Datenquelle: Statistisches Bundesamt

Zwar generieren beide Prognosen mehrere *Szenarien* für die zukünftige Entwicklung der Gesamtbevölkerung, allerdings übertrifft die Anzahl der verschiedenen Entwicklungsverläufe der Variante II aus 6.000 Prognosedurchläufen die 12 Szenarien der 12. koordinierten Bevölkerungsprognose (aus jeweils einer in sich geschlossenen deterministischen Prognosesrechnung) beträchtlich. Zudem kann in der Variante II jedem Ergebnisverlauf eine Eintrittswahrscheinlichkeit zugeordnet werden, wohingegen in der 12. koordinierten Bevölkerungsprognose alle 12 Szenarien nahezu wertungslos nebeneinander stehen. Dabei können zwar die *mittleren* Szenarien bei der Ergebnisanalyse besonders betont werden, allerdings kann dennoch weder eine Aussage zur Eintrittswahrscheinlichkeit einzelner Szenarien noch zu den Szenarien untereinander getroffen werden. Folglich können mit dem PPPM bzw. P3J nicht nur einfach weitaus mehr potentielle Entwicklungen, sondern zudem noch mit einer

Eintrittswahrscheinlichkeit gewichtet ausgegeben werden.

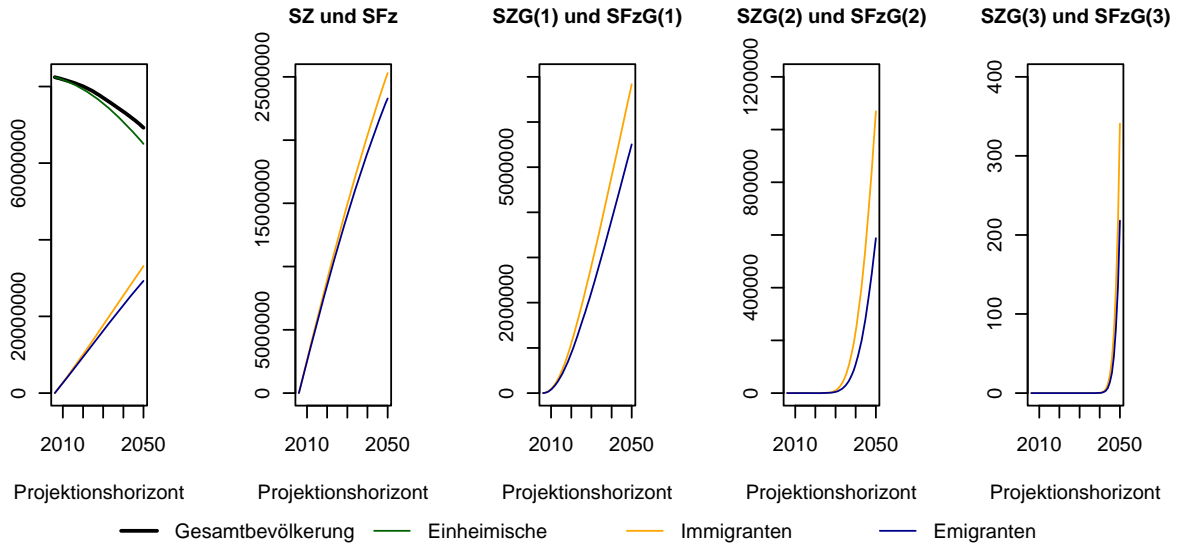


Abbildung 12.35.: Entwicklung einzelner (Sub)Populationen von 2007 bis 2050 in der Variante II der Beispielprojektion; die Prognosewerte entsprechen dem Quantil 0,5.

Neben diesem für alle probabilistischen Bevölkerungsprognosemodelle geltenden Vorzug zeichnet sich das PPPM bzw. P3J besonders durch die vielfältigen Ergebnisgrößen aus, die aus der separaten Prognose der Subpopulationen (also der Einheimischen, direkten Zu- und Fortzüge und deren Nachkommengenerationen) resultieren. So kann in der Variante II jede Subpopulation (bzw. jedes Konglomerat an Subpopulationen) aggregiert oder auch disaggregiert nach Alter und Geschlecht analysiert werden, wohingegen das bei der 12. koordinierten Bevölkerungsprognose nur für die Gesamtbevölkerung möglich ist. Die Abbildung 12.35 zeigt beispielsweise die prognostizierten Bestände für die einzelnen Subpopulationen (wie auch in der Summe für die Gesamtbevölkerung) aus der Variante II der Beispielprojektion. Dabei ist zu erkennen, dass das negative Wachstum der Einheimischen durch insgesamt mehr Immigranten als Emigranten gedämpft wird, sodass die Gesamtbevölkerung weniger stark schrumpft. Ein Blick auf die einzelnen Subpopulationen zeigt zudem, dass es in jeder Nachkommengeneration mehr Immigranten als Emigranten gibt. Darüber hinaus umfasst die Subpopulation der direkten Zuzüge in der Höhe von 25.293.861 deutlich mehr Personen als die erste, zweite und dritte Nachkommengeneration mit jeweils 6.835.869, 1.069.079 und 341 Personen im Zieljahr. Diese Entwicklung ist damit zu erklären, dass die Subpopulation

der direkten Zuzüge ab dem Startjahr durch die ständig direkt zuwandernden Personen (in nahezu allen Altern) wächst, wohingegen die Nachkommengenerationen ausschließlich durch die Geburten der vorherigen Generation wachsen und sich somit erst sukzessive von Alter null aufbauen. Diesen Aspekt verdeutlicht auch nochmal die Abbildung 12.36 mit der Alters- und Geschlechtsstruktur der ersten und zweiten Nachkommengeneration der direkten Zuzüge im Zieljahr 2050. Wächst die erste Nachkommengeneration bereits ab dem Startjahr aus den Geburten der direkten Zuzüge, kann die zweite Nachkommengeneration erst 15 Jahre später anfangen zu wachsen, da dann erst die ersten Frauen der ersten Nachkommengeneration ins reproduktive Alter kommen. Demzufolge sind in der ersten Nachkommengeneration (entsprechend der Länge des Projektionshorizonts) 44 Alter besetzt, wohingegen in der zweiten Nachkommengeneration lediglich 29 Alter besetzt sind.

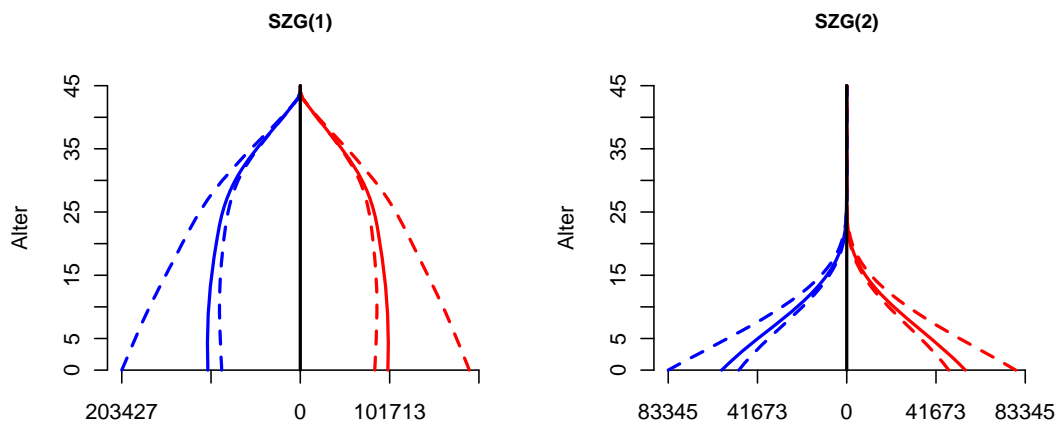


Abbildung 12.36.: Alters- und Geschlechtsstruktur der ersten (links) und zweiten (rechts) Nachkommengeneration der direkten Zuzüge im Zieljahr 2050 der Variante II der Beispielprojektion; die Prognosewerte entsprechen den Quantilen 0, 0,5 (durchgezogen) und 1.

Die genaue Prognose der Zusammensetzung der Gesamtbevölkerung aus den Einheimischen, direkten Zu- und Fortzügen und deren jeweiligen Nachkommengenerationen ist im PPPM (bzw. P3J) einzigartig und kann für diverse bevölkerungsdynamische Analysen genutzt werden. Darüber hinaus kann auch nur mit dem PPPM bzw. P3J der Einfluss einzelner Annahmen auf die Höhe der Prognoseergebnisse gesondert analysiert werden, da die Ergebniswerte mit den sie erzeugenden Annahmenverlaufskombinationen assoziiert werden. So können beispielsweise Miniprojektionen extrahiert werden, die ausschließlich

Ergebnisverläufe enthalten, die mit einem oder mehreren ausgewählten Annahmenverläufen eines oder mehrerer Modellparameter bzw. Sets entstanden sind. Für die Variante II der Beispielprojektion kann damit z. B. der Einfluss der Mortalität auf die zukünftige Entwicklung der Gesamtbevölkerung ermittelt werden, indem drei Miniprojektionen erstellt werden, die jeweils aus Ergebnisverläufen mit ausschließlich niedrigen (bzw. blauen), mittleren (bzw. roten) oder hohen (bzw. grünen) Annahmen zur Sterblichkeit für alle Subpopulationen entstanden sind; dabei zeigt sich in den Abbildungen 11.5 und 11.6 im Abschnitt 11.3.1 (ab Seite 231) deutlich, wie sich die gesamte Ergebnisverteilung der Variante II aus den Ergebnisverteilungen der grünen, roten und blauen Miniprojektion systematisch zusammensetzt; folglich kann damit auch gezeigt werden, wie die Prognoseunsicherheit in der Variante II aufgrund des Einschlusses mehrerer Annahmenverläufe in einer heterogenen Annahmenverteilung umfassender eingefangen werden kann (als in den einzelnen Miniprojektionen). Zudem können Levelplots erstellt werden, um den (kombinierten) Einfluss der Annahmenverläufe aller Modellparameter auf die Höhe der zu prognostizierenden Bestände der (Sub)Populationen zu ermitteln; einen solchen Levelplot zeigt z. B. die Abbildung 11.7 im Abschnitt 11.3.2 (ab Seite 11.3.2) für die blaue Miniprojektion.

Des Weiteren kann für eine erweiterte Ergebnisanalyse mit dem quasi-vollständigen Berechnungsmodus auch noch eine Ergebnisverteilung bestimmt werden, die aus  $n$  Ergebnisverläufen mit der höchsten Eintrittswahrscheinlichkeit besteht. Die Abbildung 12.37 stellt hierfür exemplarisch die Ergebnisverteilung für die Gesamtbevölkerung in der Variante II der Beispielprojektion nach 6.000 quasi-vollständig bzw. simulativ ermittelten Ergebnisverläufen vergleichend gegenüber. Dabei ist zu erkennen, dass die 6.000 wahrscheinlichsten Ergebnisverläufe aus der quasi-vollständigen Berechnung vollständig im unteren Bereich der Ergebnisverteilung der simulativen Berechnung enthalten sind. Allerdings ist der Unsicherheitsbereich in der quasi-vollständigen Berechnung deutlich schmaler als in der simulativen Berechnung; so deckt der Unsicherheitsbereich der quasi-vollständigen Berechnung nur 62,94 Prozent des Unsicherheitsbereichs der simulativen Berechnung ab. Dies deutet darauf hin, dass gerade die weniger wahrscheinlichen Ergebnisverläufe höhere Werte für die Gesamtbevölkerung bedingen.

**Zusammenfassung** Zusammenfassend zeigt die beispielhafte Anwendung des PPPM und P3J deren konzeptionelle Vorzüge gegenüber herkömmlichen (probabilistischen) Bevölke-

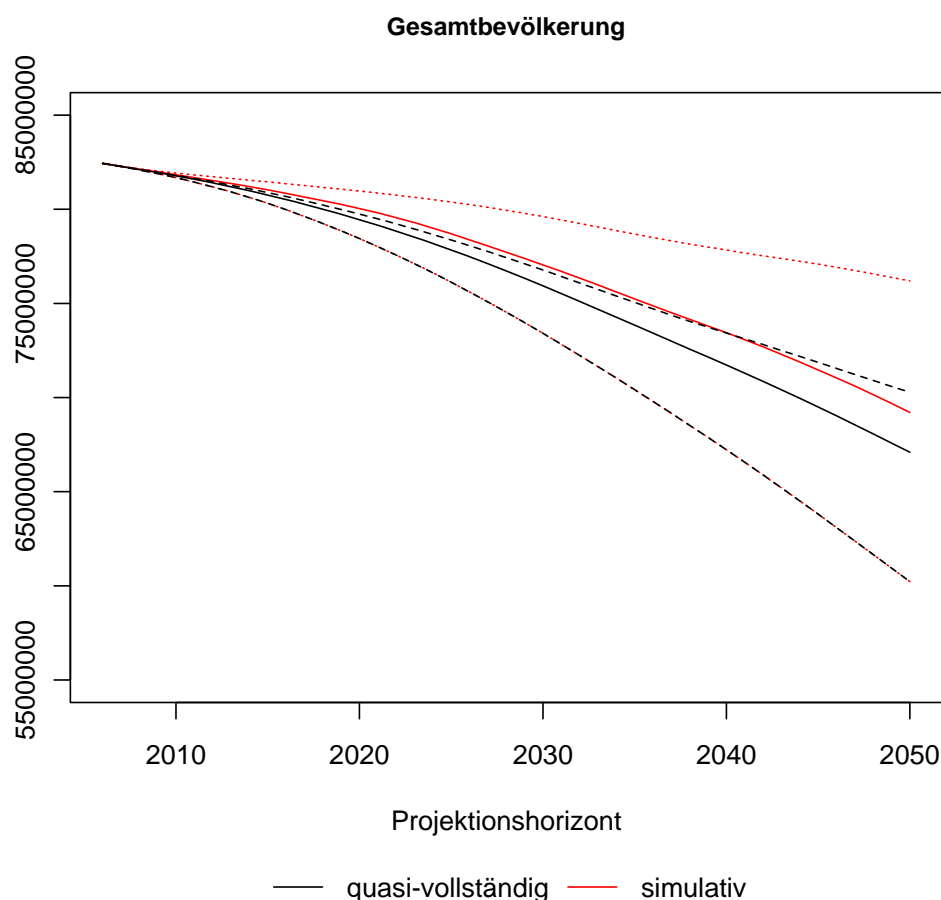


Abbildung 12.37.: Gesamtbevölkerung in der Variante II der Beispielprojektion nach 6.000 quasi-vollständig (schwarz) bzw. simulativ (rot) ermittelten Ergebnisverläufen; die Prognosewerte entsprechen den Quantilen 0, 0,5 (durchgezogen) und 1.

rungsprognose(modelle)n in nahezu allen Arbeitsschritten eine Prognoseprozesses (vgl. Abschnitt 2.2 ab Seite 18): So können mit dem PPPM und P3J — als flexible Prognoseumgebung — die Einheimischen, die direkten Zu- und Fortzüge sowie deren Nachkommengenerationen mit separaten Annahmen zur Mortalität und Fertilität probabilistisch prognostiziert werden. Dabei werden nahezu alle potentiellen zukünftigen Entwicklungen eines Modellparameters — mit verschiedenen Methoden erstellt und subjektiv gewichtet — flexibel zu einer (wenn nötig heterogenen) Annahmenverteilung verknüpft, sodass die Prognoseunsicherheit sowohl bei den Annahmen als auch bei den Ergebnissen umfassend eingefangen (und quantifiziert) werden kann. Unterstützt wird die Berechnung der Prognoseergebnisse durch die flexible



Einstellung der Korrelationsstruktur zwischen den Annahmenverläufen eines oder mehrerer Modellparameter wie auch durch die Bestimmung eines (der alternativen) Berechnungsmodi in Abhängigkeit von den allgemeinen Prognosebedingungen und den Erwartungen eines Prognostikers. Darüber hinaus können die Prognoseergebnisse vielfältig ausgewertet werden; so sind beispielsweise die Prognose der Zusammensetzung der Gesamtbevölkerung aus den Einheimischen, direkten Zu- und Fortzügen und deren jeweiligen Nachkommengenerationen wie auch die Analyse des Effekts der Annahmen auf die Höhe der Prognoseergebnisse einzigartig.

## 13. Zusammenfassung und Ausblick

### 13.1. Motivation

Bevölkerungsprognosen haben eine weitreichende Bedeutung für die zukünftige Gestaltung diverser gesellschaftlicher Lebensbereiche (vgl. Kapitel 1), sodass die fortlaufende Verbesserung ihrer Methodik unabdingbar ist. Zumal die Überprüfung vergangener Bevölkerungsprognosen (von nationalen Statistikämtern in Europa in den letzten 60 Jahren) zeigt, dass deren Genauigkeit weder durch die bessere Datenqualität noch durch die bislang verbesserte Methodik wesentlich gesteigert werden konnte [129].

Aus diesem Grund ist in dieser Arbeit das PPPM (Probabilistic Population Projection Model) zur Durchführung probabilistischer Bevölkerungsprognosen theoretisch entwickelt, implementiert und beispielhaft angewendet worden. Das PPPM soll (im Vergleich zu herkömmlichen probabilistischen Bevölkerungsprognosemodellen) die Prognoseunsicherheit *umfassender* einfangen können und dadurch die Prognosegenauigkeit erhöhen.

### 13.2. Anforderungen an ein *umfassendes* probabilistisches Bevölkerungsprognosemodell

Im ersten (theoretischen) Teil dieser Arbeit sind bestehende Ansätze zur (probabilistischen) Bevölkerungsprognose (vgl. Kapitel 3 und 4) wie auch zur Prognose einzelner Modellparameter der Mortalität (vgl. Kapitel 5), Fertilität (vgl. Kapitel 6) und Migration (vgl. Kapitel 7) betrachtet und hinsichtlich ihrer Stärken und Schwächen analysiert worden.

Diese Analyse herkömmlicher Prognoseansätze hat zahlreiche Anforderungen zur Entwicklung eines *umfassenden* Prognosemodells identifiziert, die an einzelnen Arbeitsschritten eines Prognoseprozesses (vgl. Kapitel 2) ansetzen:

1. Auswahl geeigneter Modellparameter und Bestimmung ihres Detailgrads

2. Generierung einer heterogenen Annahmenverteilung für jeden Modellparameter
3. Bestimmung einer Ergebnisverteilung für wesentliche Ergebnisgrößen
4. Erzeugung und Analyse aussagekräftiger Prognoseergebnisse

In den nachfolgenden Abschnitten 13.2.1 bis 13.2.4 wird nun jede dieser vier Anforderungen hinsichtlich ihrer allgemeinen Bedeutung und Umsetzung in *herkömmlichen* probabilistischen Ansätzen erläutert. Im Anschluss daran wird im Abschnitt 13.3 genau erläutert, inwiefern diese Anforderungen im PPPM umgesetzt worden sind.

### 13.2.1. Auswahl geeigneter Modellparameter und Bestimmung ihres Detailgrads

Ein Bevölkerungsprognosemodell ermittelt grundsätzlich aus den Annahmenwerten für die Modellparameter die Ergebniswerte für die Ergebnisgrößen (vgl. Abbildung 13.1).

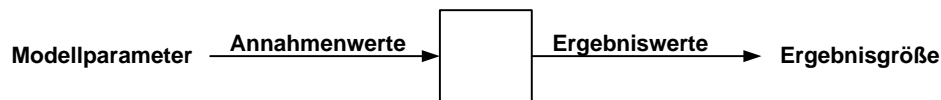


Abbildung 13.1.: Ablauf einer Bevölkerungsprognose: Für die Berechnung eines Ergebnisverlaufs pro Ergebnisgröße wird ein Annahmenverlauf pro Modellparameter benötigt.

Dabei versucht ein Prognosemodell die in der Realität auftretende Bevölkerungsbewegung möglichst wirklichkeitsgetreu abzubilden. Da jedoch nicht alle Einzelheiten bis ins letzte Detail modelliert werden können, müssen vereinfachend wesentliche Einflussfaktoren erkannt und als Modellparameter festgelegt werden (vgl. Abschnitte 4.1, 8.2 und 8.3). Demzufolge bedingt die Menge der bei einer Bevölkerungsprognose einzuschließenden Informationen die Auswahl der zu unterscheidenden Modellparameter wie auch deren Detailgrad.

Der Detailgrad der Modellparameter geht dabei auch mit der Bildung von Subpopulationen einher, für die separate Annahmen getroffen werden können. Subpopulationen sollten sich aus Personen zusammensetzen, deren demographisches Verhalten bestimmte Regelmäßigkeiten bzw. Ähnlichkeiten aufweist, sodass die Treffsicherheit der Annahmen gesteigert werden kann.

Herkömmliche (probabilistische) Makromodelle zur Bevölkerungsprognose unterscheiden meist nur die Gesamtbevölkerung nach Alter und Geschlecht. Dabei wird die einheimische Bevölkerung mit alters- und geschlechtsspezifischen Annahmen zur Fertilität und Mortalität (nach der Kohorten-Komponenten Methode (vgl. Abschnitt 3.2.1)) prognostiziert und die Nettomigranten werden (häufig erst am Ende eines Projektionsintervalls) dazu addiert (vgl. Formel 3.3); folglich werden (z. T. stark) vereinfachend dieselben Annahmen zur Fertilität und Mortalität für die Einheimischen und Nettomigranten unterstellt (vgl. Abschnitt 4.2), wodurch erhebliche Prognosefehler (vgl. Kapitel 8) entstehen können. So können herkömmliche (probabilistische) Ansätze zur Bevölkerungsprognose keine Anpassungsprozesse zwischen der Fertilität und Mortalität der Immigranten und Einheimischen (über die einzelnen Nachkommengenerationen) modellieren, obwohl diese z. B. tatsächlich in Deutschland beobachtet werden (vgl. Abschnitt 7.1.2).

#### **13.2.2. Generierung einer heterogenen Annahmenverteilung für jeden Modellparameter**

Die zukünftige Entwicklung eines Modellparameters ist mehr oder weniger mit Unsicherheit (vgl. Abschnitt 4.1) behaftet. Grundsätzlich stehen einem Prognostiker zur Erstellung einer Annahmenverteilung verschiedene Informationsquellen wie z. B. die vergangene Entwicklung, erklärende Theorien und Expertenmeinung zur Verfügung, die auch auf unterschiedliche zukünftige Entwicklungen hindeuten können. Folglich gibt es meist nicht nur eine, sondern mehrere potentielle Entwicklungen pro Modellparameter, die auch nicht nur mit einer, sondern nur mit mehreren Methoden (in verschiedenen Annahmenverläufen) eingefangen werden können (vgl. Kapitel 5, 6 und 7). Diese potentiellen Annahmenverläufe können sich dabei im Niveau, im Verlauf wie auch in der erwarteten Eintrittswahrscheinlichkeit (insbesondere bei sehr unsicheren Prognosen stark) voneinander unterscheiden.

Herkömmliche (probabilistische) Ansätze zur Bevölkerungsprognose geben nun häufig nur eine modell-basierte Methode zur Generierung der Annahmen und eine feste Verteilungsform für die Eintrittswahrscheinlichkeiten vor. Eine Annahmenverteilung entsteht dabei häufig aus zufälligen Fehlern (bzw. Abweichungen) um einen (einzig einzugebenden) mittleren Annahmenverlauf. Neben dem Niveau und dem Verlauf werden so auch die Eintrittswahrscheinlichkeiten fest vorgegeben, da die zufälligen Fehler meist normalverteilt (mit dem Mittelwert null und einer Standardabweichung) um die fixen mittleren Annahmenwerte schwanken. Die mittleren Annahmenwerte erhalten dabei immer die höchste Eintrittswahr-

scheinlichkeit (vgl. Abschnitt 4.2).

Die Vorgabe einer modell-basierten Methode zur Generierung der Annahmen und einer festen Verteilungsform für die Eintrittswahrscheinlichkeiten kann die Berücksichtigung aller potentiellen Entwicklungen eines Modellparameters, die sich aus verschiedenen Informationsquellen ergeben, so stark einschränken, dass die erzeugten Annahmenverteilungen zu homogen sind. Folglich werden dadurch die Prognoseergebnisse umso mehr verzerrt, desto unsicherer (bzw. heterogener) die zukünftige Entwicklung der einzelnen Modellparameter eigentlich ist.

### **13.2.3. Bestimmung einer Ergebnisverteilung für wesentliche Ergebnisgrößen**

Neben dem Einschluss möglichst aller potentiellen zukünftigen Entwicklungen für einen Modellparameter ist bei der Berechnung einer Ergebnisverteilung auch die Bestimmung der Korrelationsstruktur zwischen den Annahmen (eines oder auch mehrerer Modellparameter) entscheidend.

Eine Ergebnisverteilung setzt sich bei einem simulativen Bevölkerungsprognoseansatz aus verschiedenen Ergebnisverläufen zusammen; jeder Ergebnisverlauf besteht dabei aus einer Kombination von jeweils einem zufällig zu ziehenden Annahmenverlauf pro Modellparameter. Sofern die Annahmenverläufe unterschiedliche Informationsquellen bzw. Theorien widerspiegeln, hängt es von den Erwartungen eines Prognostikers ab, ob alle oder nur ein Teil aller (möglichen) Annahmenverlaufskombinationen plausibel sind. Um die Prognoseergebnisse nicht mit unplausiblen Annahmenkombinationen zu verzerren, sollte die Korrelationsstruktur zwischen den Annahmenverläufen eines oder mehrerer Modellparameter flexibel eingestellt werden können.

Herkömmliche (probabilistische) Ansätze zur Bevölkerungsprognose können mit der Festlegung bestimmter Parameter Korrelationsstrukturen zwischen den Annahmenverläufen verschiedener Modellparameter erzeugen (vgl. Abschnitt 4.2). Wenn beispielsweise eine Annahmenverteilung aus zufälligen Abweichungen um einen fest vorzugebenden mittleren Annahmenverlauf generiert wird, können diese Abweichungen sowohl über die Zeit (Autokorrelation) wie auch mit den zufälligen Abweichungen anderer Modellparameter (Kreuzkorrelation) korrelieren. So können z.B. bei dem Ansatz von Lutz et al. [174] korrelierende Annahmen für verschiedene Modellparameter, vornehmlich für die Sterblichkeit und die Fertilität (oder

aber auch bei einer mehrregionalen Prognose für einen Modellparameter über verschiedene Regionen) mit korrelierenden Zufallszahlen generiert werden (vgl. Abschnitt 4.2.4). Bei dem Ansatz von Alho und Spencer [10] kann hingegen mit den  $\kappa_x$  die Varianz der Autokorrelation,  $\delta_{x,t}$ , und der Kreuzkorrelation,  $n_x$ , für die Abweichungen von den mittleren Annahmenwerten bestimmt werden (vgl. Abschnitt 4.2.5). Generell kann jedoch die Bestimmung eines exakten Werts für die Höhe einzelner Korrelationen für einen Prognostiker schwer durchführbar sein.

#### 13.2.4. Erzeugung und Analyse aussagekräftiger Prognoseergebnisse

Wenn mit den Annahmen für die Modellparameter die Ergebniswerte für die Ergebnisgrößen berechnet worden sind, ist das eigentliche Ziel einer Bevölkerungsprognose erfüllt.

Mit hinreichend detaillierten Prognoseergebnissen können nun aber nicht nur die Ausgangsfragestellungen einer Bevölkerungsprognose beantwortet, sondern mitunter auch noch weiterführende bevölkerungsdynamische Analysen durchgeführt werden. Dabei kann beispielsweise neben der zukünftigen Entwicklung der Gesamtbevölkerung (nach Alter und Geschlecht) auch deren Zusammensetzung aus weiteren Subpopulationen (wie z.B. den Einheimischen und Personen mit Migrationshintergrund) oder auch der Einfluss einzelner Annahmen auf die Höhe der Prognoseergebnisse von Bedeutung sein.<sup>1</sup>

Herkömmliche probabilistische Makromodelle prognostizieren meist die Gesamtbevölkerung nach Alter und Geschlecht (vgl. Abschnitt 4.2), sodass weder bevölkerungsdynamische Analysen zur Zusammensetzung der Gesamtbevölkerung aus weiteren Subpopulationen (wie den Einheimischen und den Personen mit Migrationshintergrund) durchgeführt noch Auswirkungen einzelner Annahmen auf die Höhe der Prognoseergebnisse ermittelt werden können.

### 13.3. Umsetzung dieser Anforderungen im PPPM und P3J

Die herausgearbeiteten Anforderungen an ein *umfassendes* probabilistisches Bevölkerungsprognosemodell (auf der Makroebene) sind bei der Entwicklung des PPPM und seiner prototypischen Realisierung als Software, dem P3J, als wesentliche Eckpunkte verwendet worden (vgl. Kapitel 9 bis 11). Dabei dienen die Stärken herkömmlicher (fortgeschrittener)

---

<sup>1</sup>Zudem kann beispielsweise auch (mit einer Sensitivitätsanalyse) der Effekt einzelner Modellparameter auf die Prognoseergebnisse ermittelt und so das Prognosemodell kalibriert bzw. überarbeitet und weiterentwickelt werden.

probabilistischer Ansätze als Basis und die herauskristallisierten Schwächen als Ansatzpunkte zur Weiterentwicklung bzw. Verbesserung.

Nachfolgend werden nun die Besonderheiten des PPPM und P3J im Vergleich zu herkömmlichen Ansätzen beschrieben.

#### **13.3.1. Auswahl geeigneter Modellparameter und Bestimmung ihres Detailgrads**

Das PPPM ist ein probabilistisches Makromodell, dass anstatt einer Gesamtbevölkerung die Einheimischen, die direkten Zu- und Fortzüge wie auch deren Nachkommengenerationen mit separaten Annahmen zur Fertilität und Mortalität (nach Einzelalter und Geschlecht) prognostiziert. Die Anzahl der zu prognostizierenden Nachkommengenerationen kann dabei (in Abhängigkeit von der Länge des Projektionshorizonts und des Generationenabstands) frei bestimmt werden (vgl. Abschnitt 9.1).

Die separate Prognose dieser einzelnen Subpopulationen ermöglicht im Gegensatz zu herkömmlichen probabilistischen Ansätzen die Berücksichtigung von Brutto- statt Nettomigration wie auch die Annahme eines unterschiedlichen demographischen Verhaltens zwischen Personen mit Migrationshintergrund und den Einheimischen. So kann beispielsweise ein sukzessiver Anpassungsprozess von der höheren Fertilität und der niedrigeren Mortalität der direkten Zuzüge an das Niveau der Einheimischen über die einzelnen Nachkommengenerationen modelliert werden. Exemplarisch ist dies auch in der umfassenden Beispielprojektion von 2007 bis 2050 mit realen Daten für Deutschland gezeigt worden (vgl. die Abschnitte 12.4.1 und 12.4.2).

#### **13.3.2. Generierung einer heterogenen Annahmenverteilung für jeden Modellparameter**

Im PPPM können außerdem unbegrenzt viele, extern mit verschiedenen Methoden zu generierende Annahmenverläufe für jeden Modellparameter eingegeben *und* mit frei bestimmbaren Eintrittswahrscheinlichkeiten subjektiv gewichtet werden, sodass (im Gegensatz zu herkömmlichen probabilistischen Bevölkerungsprognoseansätzen) auch bei besonders unsicheren Bevölkerungsprognosen alle potentiellen Entwicklungen in einer heterogenen Annahmenverteilung umfassend und flexibel berücksichtigt werden können (vgl. Abschnitt

10.2.1). So werden auch in der umfassenden Beispielprojektion für (nahezu) jeden Modellparameter heterogene Annahmenverteilungen erstellt, wobei die einzelnen Annahmenverläufe z. T. mit unterschiedlichen Methoden erstellt und zudem mit der Zuordnung von Eintrittswahrscheinlichkeiten subjektiv gewichtet werden (vgl. Abschnitt 12.4).

Die Erstellung einer heterogenen Annahmenverteilung kann im PPPM noch durch die Option, repräsentative Annahmenverläufe mit stochastischer Variation für die Modellparameter einzugeben, gesteigert werden (vgl. Abschnitt 10.2.1). Ein repräsentativer Annahmenverlauf gibt dabei das Niveau wie auch den Verlauf der Annahmenwerte richtungsweisend vor. In einem Prognosedurchlauf werden dann diese repräsentativen Annahmenwerte mit einer frei einstellbaren Standardabweichung zufällig (über das Alter oder/und die Zeit) variiert, wobei jeweils die repräsentative Altersstruktur aus Plausibilitätsgründen beibehalten wird. Im PPPM können (im Gegensatz zu herkömmlichen probabilistischen Bevölkerungsprognosemodellen) für jeden Modellparameter anstatt nur eines *mehrere* repräsentative Annahmenverläufe eingegeben werden, sodass das jeweilige Spektrum potentieller zukünftiger Entwicklungen noch dichter bzw. breiter abgedeckt werden kann. Dieser Effekt wird in der Beispielprojektion durch den Vergleich der Varianten III und IV sowohl für die Annahmen (vgl. die Abbildungen 12.26 und 12.27) als auch für die Ergebnisse (vgl. die Abbildungen 12.31 und 12.32) gezeigt.

Daneben können die repräsentativen Annahmenverläufe mit stochastischer Variation auch dazu genutzt werden, eine probabilistische Bevölkerungsprognose mit dem PPPM durchzuführen, obwohl ein Prognostiker für die Modellparameter immer nur einen Annahmenverlauf generiert. Folglich können die repräsentativen Annahmenverläufe bei weniger unsichereren Bevölkerungsprognosen mit relativ homogenen Annahmenverteilungen auch eine Reduktion des Arbeitsaufwands bei der Annahmengenerierung bedeuten.

#### **13.3.3. Bestimmung einer Ergebnisverteilung für wesentliche Ergebnisgrößen**

Im PPPM kann zudem die Prognosegenauigkeit durch die flexible Einstellung der Korrelation zwischen den Annahmenverläufen eines oder mehrerer Modellparameter gesteigert werden (vgl. Abschnitt 10.2.2). Im *Open Type* des PPPM sind die Annahmenverläufe aller Modellparameter uneingeschränkt kombinierbar, sodass z. T. unplausible Annahmenkombinationen, die nicht den Erwartungen des jeweiligen Prognostikers entsprechen, zufällig gezogen



werden. Im *Limited Type* können hingegen diese unplausiblen Kombinationen durch die Einführung von Settypen und Sets gezielt eliminiert werden, indem die Modellparameter und Annahmenverläufe, zwischen denen der jeweilige Prognostiker eine (plausible) Korrelation vermutet, flexibel bestimmt und zusammengeschlossen werden. In den Settypen werden dazu die Modellparameter und in den Sets die entsprechenden Annahmenverläufe miteinander verbunden. Die Elimination unplausibler Kombinationen führt somit zu einer Bereinigung einer möglicherweise durch unplausible Kombinationen verzerrten Ergebnisverteilung, sodass dadurch die Prognosegenauigkeit gesteigert werden kann. Dieser Effekt wird in der Beispielprojektion durch den Vergleich der Varianten I und II (z.B. in den Abbildungen 12.31 und 12.32) demonstriert.

Darüber hinaus ist der Berechnungsprozess im PPPM gegenüber herkömmlichen probabilistischen Bevölkerungsprognosemodellen äußerst flexibel, da die simulative, vollständige oder/und quasi-vollständige Berechnungsvariante (in Abhängigkeit von der Anzahl der einzugebenden Annahmenverläufe) optional eingestellt werden kann (vgl. Abschnitt 10.3.2). Die vollständige Berechnungsvariante ist bei weniger unsicheren Bevölkerungsprognosen mit relativ wenigen Annahmenverläufen pro Modellparameter vorteilhaft, da sie eine unveränderliche Ergebnisverteilung erzeugt, bei der jeder zu unterscheidende Ergebnisverlauf nur einmal berechnet und entsprechend seiner Eintrittswahrscheinlichkeit berücksichtigt wird. Sofern nun aber relativ viele Annahmenverläufe pro Modellparameter in einer unsicheren Bevölkerungsprognose berücksichtigt werden, steigt die Anzahl der zu unterscheidenden Annahmenverlaufskombinationen bzw. Ergebnisverläufe dermaßen, dass der Rechenaufwand bei der vollständigen im Vergleich zur simulativen Berechnungsvariante unverhältnismäßig groß wird bzw. gar nicht mehr zu bewältigen ist (vgl. Abschnitt 10.3.1). Die simulative Berechnung kann die Ergebnisverteilung trotz dieser Komplexität (mit deutlich weniger Prognosedurchläufen als zu unterscheidenden Ergebnisverläufen) approximativ schätzen, wohingegen die quasi-vollständige Berechnung vereinfachend nur einen Anteil der wahrscheinlichsten Ergebnisverläufe angibt. Damit schließt die quasi-vollständige Berechnung die unwahrscheinlicheren Ergebnisverläufe aus, sodass sie das zukünftige Entwicklungsspektrum (unsicherer Bevölkerungsprognosen mit heterogenen Annahmenverteilungen) im Gegensatz zur simulativen Berechnung nur unvollständig abdecken kann. Folglich sollte bei sicheren bzw. unsicheren Bevölkerungsprognosen mit relativ wenigen bzw. vielen Annahmenverläufen die vollständige bzw. simulative Berechnungsvariante verwendet werden. Die quasi-vollständige Berechnungsvariante kann hingegen ergänzend zur erweiterten Analyse

der Prognoseergebnisse eingesetzt werden, indem sie beispielsweise zeigt, welcher Teil einer simulativ zu erzeugenden Ergebnisverteilung durch die  $x$  wahrscheinlichsten Ergebnisverläufe abgedeckt wird.

In der Beispielprojektion ist jede der vier Varianten aufgrund der heterogenen Annahmenverteilungen für nahezu jeden Modellparameter simulativ (anstatt vollständig) berechnet worden (vgl. Abschnitt 7.1.2). Zudem ist die Variante II aber auch noch für eine erweiterte Ergebnisanalyse quasi-vollständig ausgeführt worden (vgl. Abbildung 12.37).

#### 13.3.4. Erzeugung und Analyse aussagekräftiger Prognoseergebnisse

Mit dem PPPM können relativ komplexe Bevölkerungsprognosen durchgeführt werden, in denen zahlreiche Annahmenverläufe für jede der einzeln zu prognostizierenden Subpopulationen berücksichtigt werden können. Der komplexe Aufbau des PPPM ermöglicht dabei die Generierung vielfältiger Prognoseergebnisse, die in ihrer Aussagekraft die Prognoseergebnisse herkömmlicher probabilistischer Ansätze übertreffen. So können die grundlegenden Prognoseergebnisse des PPPM, also die prognostizierten Jahresend- und Jahresdurchschnittsbestände, für alle (separat zu prognostizierenden und teilweise daraus zu aggregierenden) (Sub)Populationen nach Einzelalter und Geschlecht (probabilistisch) ausgewertet werden. Außerdem können diverse weitere Ergebnisgrößen wie z. B. demographische Alterungsmaße aus den grundlegenden Prognoseergebnissen zusätzlich berechnet und ausgegeben werden. Neben der Berechnung vielfältiger Ergebnisgrößen diverser (Sub)Populationen, kann die Ergebnisanalyse durch die Assoziation der Ergebniswerte mit den sie erzeugenden Annahmenverläufen erweitert werden, sodass nicht nur die Prognoseergebnisse für sich, sondern in Abhängigkeit der zufällig gezogenen Annahmenverläufe analysiert werden können. So können Ergebnisverläufe extrahiert und in so genannten *Miniprojektionen* zusammengeschlossen werden, die alle einen oder mehrere zufällig gezogene Annahmenverläufe gemein haben. Ergänzend dazu können auch noch Levelplots erstellt werden, die die zufällig gezogenen Annahmenverläufe pro Modellparameter für alle aufsteigend sortierten Ergebnisverläufe einer beliebigen Ergebnisgröße darstellen. Sowohl die Miniprojektionen als auch die Levelplots erlauben somit einen Rückschluss von der Höhe der Prognosewerte auf die zufällig gezogenen Annahmenverläufe (und umgekehrt) (vgl. Abschnitt 11.3.1).

Im Gegensatz zu herkömmlichen probabilistischen Bevölkerungsprognoseansätzen kann so z. B. mit dem PPPM bzw. P3J die Zusammensetzung der in der Beispielprojektion prognos-

tizierten Gesamtbevölkerung aus den einzelnen Subpopulationen (vgl. Abbildung 12.35) oder auch der Einfluss der einzelnen Annahmen auf die Höhe der Gesamtbevölkerung in entsprechenden Miniprojektionen (vgl. die Abbildungen 11.5 und 11.6) und Levelplots (vgl. Abbildung 11.7) analysiert werden. Aufgrund der Vielzahl der Prognoseergebnisse des PPPM kann deren automatische Generierung für einen Nutzer vorteilhaft sein; deshalb ist die automatische Erstellung eines Ergebnisreports für die grundlegenden Prognoseergebnisse des PPPM bereits in der prototypischen Realisierung des PPPM als Software, dem P3J, umgesetzt (vgl. Abschnitt 11.2).

### 13.3.5. Prototypische Realisierung des PPPM als Software im P3J

Um die konzeptionellen Besonderheiten des PPPM praktisch umzusetzen, ist es auf der Basis des quelloffenen Modellierungs- und Simulationsrahmenwerks JAMES II implementiert worden (vgl. Abschnitte 9.3, 10.4 und 11.4). Die daraus resultierende Software, P3J, ist wie JAMES II in der weit verbreiteten Programmiersprache Java realisiert worden und zeichnet sich durch eine hohe Flexibilität gegenüber den technischen Anforderungen eines Nutzers (bezüglich der vorhandenen Hard- und Software) aus. So ist es möglich, die Berechnung einer Prognose mit Hilfe eines Mehrkernprozessors zu beschleunigen und die verwendeten bzw. erzeugten Daten in verschiedensten Datenbanken abzulegen. Durch das von JAMES II bereitgestellte *Plug'n Simulate* Konzept ist es einfach möglich, die entwickelten Komponenten zu ergänzen und zu erweitern; so können dadurch beispielsweise alternative Berechnungsvarianten — wie die simulative und (quasi-)vollständige Ausführung — nahtlos in die entwickelte Software integriert werden. Darüber hinaus stellt das P3J eine graphische Bedienoberfläche zur Verfügung, die die Eingabe und Verwaltung der Annahmenverläufe vereinfacht. Nach der Ausführung einer Bevölkerungsprognose mit dem P3J kann der Nutzer die Prognoseergebnisse als Rohdaten (in Form von komma-separierten Dateien) exportieren oder automatisch mit Hilfe von R und L<sup>A</sup>T<sub>E</sub>X in einem Ergebnisreport zusammenfassen lassen.

## 13.4. Ausblick bzw. noch umzusetzende Anforderungen im PPPM und P3J

Obgleich im PPPM und P3J wesentliche Fortschritte gegenüber herkömmlichen probabilistischen Bevölkerungsprognoseansätzen erzielt worden sind, gibt es noch weitere Ansatzpunkte zur Weiterentwicklung, die nachfolgend kurz diskutiert werden sollen.

### 13.4.1. Auswahl geeigneter Modellparameter und Bestimmung ihres Detailgrads

#### Zu unterscheidende Subpopulationen

Die Aufteilung der Gesamtbevölkerung im PPPM ermöglicht zwar im Gegensatz zu herkömmlichen Bevölkerungsprognosemodellen die Simulation sukzessiver Anpassungsprozesse im demographischen Verhalten zwischen den direkten Zuzügen und den Einheimischen über die nachfolgenden Nachkommengenerationen. Allerdings stellen weder die direkten Zu- und Fortzüge noch deren einzelne Nachkommengenerationen Subpopulationen mit homogenem demographischen Verhalten dar, sodass sich auch innerhalb dieser Subpopulationen die Fertilität und Mortalität unterscheiden kann. So können z. B. bei den direkten Zuzügen Fertilitäts- und Mortalitätsunterschiede zwischen Personen unterschiedlicher Herkunft auftreten. Folglich wäre eine zusätzliche Unterteilung der Migranten-Subpopulationen nach ihrer Herkunft realistisch. Jedoch würden sich aus dieser realistischen Modellierung auch etliche Probleme ergeben: Erstens würde die Komplexität des PPPM deutlich zunehmen, da dann nicht nur eine Annahmenverteilung für jede Migranten-Subpopulation, sondern eine Annahmenverteilung für jede Personengruppe derselben Herkunft in jeder Migranten-Subpopulation erzeugt werden müsste. Zweitens müsste das Merkmal Herkunft genau definiert werden, sodass klar würde, nach welchen Gebietseinheiten (z. B. Kontinente oder Länder) die Personen eingeteilt werden würden. Je kleiner dabei die gewählten Gebietseinheiten sein würden, desto mehr Herkunftsregionen würden unterschieden werden und desto schwieriger wäre der Bezug entsprechender demographischer Daten zur Fertilität und Mortalität nach Einzelalter und Geschlecht.

Selbst bei der tatsächlichen Aufteilung der Gesamtbevölkerung in Einheimische, direkte Zu- und Fortzüge und deren einzelne Nachkommengenerationen im PPPM müssen z. T. Ausgangsdaten zur Fertilität und Mortalität für einzelne Migranten-Subpopulationen geschätzt werden (vgl. Abschnitte 12.4.1 und 12.4.2). Allerdings wäre der Prognosefehler, insbesondere bei Langzeitprognosen, derart groß, wenn Netto- statt Bruttomigration mit einheitlichen statt separaten Annahmen zur Fertilität und Mortalität in einer Bevölkerungsprognose berücksichtigt würde, dass der bereits (in Deutschland) begonnene Prozess der zunehmenden Erfassung und Analyse derlei Daten (auch aus diesem Grund) fortgeführt werden sollte.

Daneben kann auch die Einteilung der Migranten in direkte Zu- und Fortzüge und deren

einzelne Nachkommengenerationen insofern zu einer Verzerrung führen, alsdass die jeweils getroffenen Fertilitäts- und Mortalitätsannahmen nicht auf alle Personen einer Migranten-Subpopulation gleichermaßen zutreffen. So wachsen z. B. die mitwandernden (Klein)Kinder der direkten Zuzüge weitestgehend im Zielland auf, sodass für sie quasi die Fertilitäts- und Mortalitätsannahmen der ersten Nachkommengeneration zutreffender wären als die der direkten Zuzüge. Dies kann allerdings mit dem derzeitigen Modellaufbau des PPPM nicht beachtet werden, sodass das PPPM an dieser Stelle entsprechend modifiziert und weiterentwickelt werden sollte. Die Prognoseergebnisse sollten von diesem Effekt allerdings nur marginal verändert werden; generell gilt, je stärker sich die Annahmen zur Fertilität und Mortalität zwischen den direkten Zu- und Fortzügen und ihrer jeweiligen ersten Nachkommengeneration unterscheiden und je größer der Anteil der (Klein)Kinder an der jeweiligen Subpopulation ist, desto größer wird sich dieser (verzerrende) Effekt auf die Prognoseergebnisse auswirken.

#### **Modellparameter *direkte Fortzüge* — $Fz^m$ und $Fz^w$**

Eine weitere Umgestaltung des PPPM betrifft die Modellparameter der direkten männlichen und weiblichen Fortzüge. Im PPPM bestehen die Annahmen zur Migration vereinfachend aus totalen Immigrations- und Emigrationszahlen nach Alter und Geschlecht. Wenn anstatt der Emigrationszahlen Emigrationsraten verwendet werden würden, könnten die Emigranten fortlaufend auf die zu prognostizierende Bevölkerung abgestimmt werden. Im Vergleich dazu muss im PPPM schon am Prognosebeginn die Anzahl der Emigranten für jedes Projektionsintervall bestimmt werden. Für die Immigranten wäre der Ersatz der Immigrationszahlen mit Immigrationsraten jedoch nicht von Vorteil, da dann die entsprechende Risikobevölkerung der Weltbevölkerung mit Ausnahme der Zielbevölkerung entspricht. Der Einsatz von Bestandszahlen für die Immigranten und Emigranten ermöglicht zwar dieselbe Vorgehensweise bei der Annahmengenerierung, allerdings könnte der Einsatz von Emigrationsraten vorteilhaft sein.

#### **13.4.2. Generierung einer heterogenen Annahmenverteilung für jeden Modellparameter**

##### **Entwicklung eines Annahmengenerierungstools**

Die Trennung der Annahmengenerierung, der Berechnung der Prognoseergebnisse und der Ergebnisanalyse ermöglicht eine hohe Flexibilität in jedem Arbeitsschritt einer Bevölkerungsprognose mit dem PPPM. Dennoch sollte eine Weiterentwicklung des PPPM die Erstellung eines Annahmengenerierungstools vorsehen, mit dem der Prognostiker bei der Annahmengenerierung

nerierung unterstützt werden kann. Häufig verwendete bzw. besonders geeignete (statistische) Methoden können dann zur Generierung der Annahmenverläufe für die einzelnen Modellparameter des PPPM zur Verfügung gestellt werden. Insbesondere bayesianische Methoden könnten die Annahmengenerierung für das PPPM unterstützen, da sie Informationen aus Beobachtungsdaten (likelihood) mit korrigierender Expertenmeinung (prior) verbinden und damit der grundlegenden Konzeption des PPPM besonders entsprechen.

#### **Subjektive Bestimmung der Eintrittswahrscheinlichkeiten**

Im PPPM können für jeden Modellparameter unbegrenzt viele, extern mit verschiedenen Methoden zu generierende Annahmenverläufe eingegeben und mit subjektiv bestimmbar Eintrittswahrscheinlichkeiten gewichtet werden, um möglichst alle potentiellen zukünftigen Entwicklungen in einer heterogenen Annahmenverteilung einzufangen. Dabei sollten die Eintrittswahrscheinlichkeiten die Erwartungen des Prognostikers darüber wiedergeben, wie wahrscheinlich er (oder sie) ein Eintreten des jeweiligen Annahmenverlaufs einschätzt. In einer weiterführenden Arbeit könnte die Bestimmung dieser Eintrittswahrscheinlichkeiten genauer thematisiert werden, sodass letztendlich der Einsatz eines oder mehrerer Experten für ein spezielles oder alle Fachgebiete (Fertilität, Mortalität, Migration) auf der Basis entsprechender Tests empfohlen werden kann.<sup>2</sup>

#### **13.4.3. Bestimmung einer Ergebnisverteilung für wesentliche Ergebnisgrößen**

##### **Bestimmung der Anzahl der Prognosedurchläufe**

Es sollte außerdem noch eine Methode zur Bestimmung der Anzahl der zu benötigten Prognosedurchläufe gefunden werden, damit die jeweils resultierenden Ergebnisverteilungen auch bei mehreren simulativen Ausführungen kaum noch voneinander abweichen. Dabei sind die Heterogenität der Annahmenverteilungen hinsichtlich der Annahmenwerte und Eintrittswahrscheinlichkeiten wie auch die Anzahl aller möglichen Annahmenverlaufskombinationen entscheidend. Die Anzahl der Prognosedurchläufe sollte steigen, desto mehr Annahmenverlaufskombinationen möglich, desto niedriger deren Eintrittswahrscheinlichkeiten und desto heterogener die Annahmenwerte sind. Hierfür können bereits etablierte statistische Metho-

---

<sup>2</sup>Beispielsweise könnte in diesem Zusammenhang die Delphi-Methode [13] zur Generierung einzelner Annahmenverläufe eingesetzt werden. Genauere Informationen dazu sind z. B. auch unter <http://armstrong.wharton.upenn.edu/delphi2/> zu finden.

den aus den Bereichen der Optimierung und Simulation [88] verwendet und entsprechend umgewandelt bzw. erweitert werden.

#### **Effekt *eines* im Vergleich zu *mehreren* (repräsentativen) Annahmenverläufen pro Modellparameter auf die Ergebnisverteilung**

In der Beispielprojektion ist durch den Vergleich der Varianten III und IV getestet worden, inwieweit *ein* repräsentativer Annahmenverlauf pro Modellparameter (in der Variante III) die Ergebnisverteilung mit *mehreren* repräsentativen Annahmenverläufen pro Modellparameter (in der Variante IV) wiedergeben kann (vgl. Abschnitt 12.5). Dabei ist herausgekommen, dass der Ersatz aller mit immer nur einem repräsentativen Annahmenverlauf pro Modellparameter zu einem Ausschluss zahlreicher potentieller Zukunftsszenarien führt, sodass die eigentlich heterogene Annahmen- und Ergebnisverteilung der Variante II nur unvollständig in der Variante III wiedergegeben werden kann. Dieser Test könnte ergänzt werden, indem das Ausmaß der stochastischen Variation für die repräsentativen Annahmenverläufe in der Variante III systematisch verändert und die jeweils entstehende Ergebnisverteilung mit der aus der Variante IV verglichen wird.

#### **13.4.4. Realisierung der potentiellen PPPM-Weiterentwicklungen im P3J**

Die in den Abschnitten 13.4.1 bis 13.4.3 diskutierten Erweiterungsmöglichkeiten bezüglich des PPPM sollten auch in der Weiterentwicklung des P3J berücksichtigt werden. Methodische Verfeinerungen des Modells, etwa durch zusätzliche Subpopulationen oder der Verwendung von Emigrationsraten, müssen sowohl in der Modellstruktur als auch in den Berechnungsalgorithmen reflektiert werden. Weitere Werkzeuge zur Unterstützung des Nutzers, wie das in Abschnitt 13.4.2 motivierte Annahmengenerierungstool, können über das *Plug'n Simulate* Konzept von JAMES II einfach in die bestehenden Strukturen integriert werden. Die Entwicklung solcher Hilfsmittel setzt allerdings voraus, dass die Nutzer über eine geeignete Dokumentation, beispielsweise in Form eines Handbuchs, über die vielfältigen Funktionen des P3J informiert werden. Dies gilt auch für die Verwendung von spezialisierten statistischen Verfahren, wie sie in Abschnitt 13.4.3 zur Abschätzung der benötigten Prognosedurchläufe vorgeschlagen werden.

Um die tatsächlich benötigte Anzahl an Prognosedurchläufen zur Ermittlung eines signifikanten Unsicherheitsbereichs festzustellen, kann auf bereits vorhandene JAMES II-Komponenten zurückgegriffen werden, da diese auch für andere Anwendungsbereiche

stochastischer Simulationen relevant sind. Diese Komponenten kontrollieren die Anzahl der Prognosedurchläufe, indem sie die bisher erzielten Ergebnisse statistisch auswerten und im Falle einer nicht zufrieden stellenden Signifikanz die Ausführung von weiteren Durchläufen veranlassen. Durch die Wiederverwendung von bereits entwickelten Komponenten kann die Entwicklungszeit, die für das P3J benötigt wird, reduziert werden. Zudem sind diese Komponenten bereits in anderen Kontexten getestet worden, wodurch sich ihre Validität erhöht.

Abgesehen von den konzeptionellen Erweiterungen des PPPM, die im P3J umgesetzt werden müssten, gibt es auch noch einige praktische Aspekte mit Verbesserungspotential. Obwohl die aktuelle Version an sich benutzbar ist, sollte sie als *Prototyp* angesehen werden; sie ist beispielsweise weder auf Geschwindigkeit optimiert, noch ist der Installationsprozess (auf verschiedenen Plattformen) ausreichend getestet. Neben einem Handbuch ist auch die Erstellung einer Online-Hilfe wünschenswert, sowie eine weitere Vereinfachung der Eingabemasken. Beispielsweise könnten fehlerhafte Eingaben, wie eine Eintrittswahrscheinlichkeit größer als 1, farblich hervorgehoben werden.

## 13.5. Schlussbemerkung

Grundsätzlich werden mit dem PPPM (menschliche) Bevölkerungen nach der Zahl und Struktur *umfassend* prognostiziert. Wesentliche Strukturvariablen stellen dabei das Alter, das Geschlecht wie auch die Zugehörigkeit zu einer der zu unterscheidenden Subpopulationen dar.

Folglich kann das PPPM Basisdaten zur zukünftigen Entwicklung einer Bevölkerung bereitstellen, die für diverse gesellschaftliche Bereiche — im sozialen, ökonomischen und ökologischen Kontext — von grundlegender Bedeutung sind.

Die neuartige Konzeption des PPPM erlaubt dabei (im Vergleich zu herkömmlichen (probabilistischen) Bevölkerungsprognosemodellen) die besonders flexible und genaue Prognose einer Bevölkerung auf der Makroebene. *Genau* bedeutet in diesem Fall jedoch nicht *eine* Lösung bzw. Bevölkerungsentwicklung, die mit großer Wahrscheinlichkeit falsch ist, sondern eine *approximative* Lösung bzw. Ergebnisverteilung, die mit großer Wahrscheinlichkeit die tatsächlich eintretende Bevölkerungsentwicklung enthält. Im Sinne der Aussagen Feynmans (vgl. Kapitel 1) und Tukeys wäre es demnach falsch zu fragen, wie sich eine Bevölkerung in



Zukunft konkret entwickeln wird; vielmehr sollte gefragt werden, wie sich eine Bevölkerung — zerlegt in Subpopulationen mit annähernd homogenem demographischen Verhalten — *wahrscheinlich* in Zukunft entwickeln wird.

*[...] Far better an approximate answer to the right question, than the exact answer to the wrong question, which can always be made precise [...]*

John W. Tukey

Um die *wahre* und *wahrscheinliche* Annahmen- und schließlich Ergebnisverteilung überhaupt *umfassend* mit (nahezu) allen potentiellen zukünftigen Entwicklungen abdecken zu können, werden nicht nur *eine*, sondern *mehrere* potentielle Entwicklungen für jeden Modellparameter in einer Bevölkerungsprognose mit dem PPPM berücksichtigt. Folglich stellen das PPPM und P3J eine *umfassende* und flexible Prognoseumgebung dar, mit der ein Prognostiker in Abhängigkeit der jeweiligen Prognosebedingungen und (subjektiven) Erwartungen wesentliche Aspekte, wie z.B. die einzusetzenden Informationsquellen und Methoden zur Annahmengenerierung und Berechnung der Ergebnisse, selbst bestimmen kann.

Letztlich verwirklichen das PPPM und P3J mit dieser Vorgehensweise die Überzeugung, dass die Unsicherheit der zukünftigen Entwicklung (der Modellparameter und Ergebnisgrößen) nur *umfassend* eingefangen bzw. abgedeckt werden kann, wenn der Prognostiker von zwar auch unterstützenden, aber z.T. auch (stark) einschränkenden modell-basierten Voreinstellungen (herkömmlicher (probabilistischer) Bevölkerungsprognosemodelle) befreit wird. Dies gelingt allerdings nur, wenn dem Prognostiker im Gegenzug auch wieder mehr Entscheidungsfreiheit und Verantwortung bei der Durchführung einer Bevölkerungsprognose übertragen wird.

# Literaturverzeichnis

- [1] ABBASI-SHAVAZI, Mohammad J. ; McDONALD, Peter: Fertility and Multiculturalism: Immigrant Fertility in Australia, 1977-1991. In: *International Migration Review of Economics and Statistics* 34 (2000), S. 215–242
- [2] AHLBURG, Dennis A.: Simple versus Complex Models: Evaluation, Accuracy, and Combining. In: *Mathematical Population Studies* 5 (1995), Nr. 3, S. 281–290
- [3] ALDERS, Maarten ; DE BEER, Joop: Assumptions on Fertility in Stochastic Population Forecasts. In: *International Statistical Review* 72 (2004), Nr. 1, S. 65–79
- [4] ALDERS, Maarten ; KEILMAN, Nico ; CRUIJSEN, Harri: Assumptions for long-term stochastic population forecasts in 18 European countries. In: *European Journal of Population and Development Review* 23 (2007), Nr. 1, S. 33–69
- [5] ALHO, Juha M.: Effect of aggregation on the estimation of trend in mortality. In: *Mathematical Population Studies* 3 (1991), S. 53–67
- [6] ALHO, Juha M.: *Katsauksia Översikter Reviews*. Bd. 4: *A Stochastic Forecast of the Population of Finland*. Statistics Finland, 1998
- [7] ALHO, Juha M. ; ALDERS, Maarten ; CRUIJSEN, Harri ; KEILMAN, Nico ; NIKANDER, Timo ; PHAM, Dinh Q.: New forecast: Population decline postponed in Europe. In: *Statistical Journal of the United Nations ECE* 23 (2006), 1-10. [folk.uio.no/keilman/UNECE.pdf](http://folk.uio.no/keilman/UNECE.pdf). – Zugriff 04/2011
- [8] ALHO, Juha M. ; CRUIJSEN, Harri ; KEILMAN, Nico: Empirically based specification of forecast uncertainty. In: ALHO, Juha M. (Hrsg.) ; JENSEN, Svend E. H. (Hrsg.) ; LASSILA, Jukka (Hrsg.): *Uncertain Demographics and Fiscal Sustainability*. Cambridge University Press, 2008, S. 34–54

- [9] ALHO, Juha M. ; SPENCER, Bruce D.: The Practical Specification of the Expected Error of Population Forecasts. In: *Journal of Official Statistics* 13 (1997), Nr. 3, S. 203–225
- [10] ALHO, Juha M. ; SPENCER, Bruce D.: *Statistical Demography and Forecasting*. Springer Science+Business Media, Inc., 2005
- [11] ANDREEV, Kirill F. ; VAUPEL, James W.: Forecasts of Cohort Mortality after Age 50. In: *MPIDR Working Paper WP 2006-012* (2006). [www.demogr.mpg.de/papers/working/wp-2006-012.pdf](http://www.demogr.mpg.de/papers/working/wp-2006-012.pdf). – Zugriff 04/2011
- [12] ARMSTRONG, J. S.: Evaluating Forecasting Methods. In: ARMSTRONG, J. S. (Hrsg.): *Principles of Forecasting*, Kluwer Academic Publishers, 2001, S. 441–472
- [13] ARMSTRONG, J. S.: *Principles of Forecasting*. Kluwer Academic Publishers, 2001
- [14] ARMSTRONG, J. S. ; COLLOPY, Fred: Error Measures for Generalizing about Forecasting Methods: Empirical Comparisons. In: *International Journal of Forecasting* 8 (1992), S. 69–80
- [15] ARTHUR, W. B. ; ESPENSHADE, Thomas J.: Immigration Policy and Immigrants' Ages. In: *Population and Development Review* 14 (1988), Nr. 2, S. 315–326
- [16] BANKS, Jerry ; CARSON, II, John S. ; NELSON, Barry L. ; NICOL, David M. ; FABRYCKY, W. J. (Hrsg.) ; MIZE, J. H. (Hrsg.): *Discrete-Event System Simulation. Third Edition*. Prentice Hall International Series In Industrial And Systems Engineering, 2001
- [17] BEAN, Frank D. ; CULLEN, Ruth M. ; STEPHEN, Elizabeth H. ; SWICEGOOD, C. G.: Generational differences in fertility among mexican Americans: implications for assessing the effects of immigration. In: *Social Science Quarterly* 65 (1984), Nr. 2, S. 573–582
- [18] BELL, William R.: Comparing and Assessing Time Series Methods for Forecasting Age-Specific Fertility and Mortality Rates. In: *Journal of Official Statistics* 13 (1997), Nr. 3, S. 279–303
- [19] BENJAMIN, B. ; SOLIMAN, A. S.: *Mortality on the Move. Methods of mortality projection*. Actuarial Education Service, 1993
- [20] BERNADELLI, H.: Population waves. In: *Journal of the Burma Research Society* 31 (1941), S. 1–18

- [21] BIJAK, Jakub: Forecasting International Migration: Selected Theories, Models, And Methods. In: *CEFMR Working Paper 4/2006* (2006). [http://www.cefmr.pan.pl/docs/cefmr\\_wp\\_2006-04.pdf](http://www.cefmr.pan.pl/docs/cefmr_wp_2006-04.pdf). – Zugriff 04/2011
- [22] BIJAK, Jakub: Bayesian methods in international migration forecasting. In: RAYMER, James (Hrsg.) ; WILLEKENS, Frans (Hrsg.): *International Migration in Europe. Data, Models and Estimates*. John Wiley & Sons Ltd., 2008, S. 255–281
- [23] BIJAK, Jakub ; KUPISZEWSKI, Marek ; KICINGER, Anna: International migration scenarios for 27 European countries, 2002-2052. In: *CEFMR Working Paper 4/2004* 4 (2004). [http://www.cefmr.pan.pl/docs/cefmr\\_wp\\_2004-04.pdf](http://www.cefmr.pan.pl/docs/cefmr_wp_2004-04.pdf). – Zugriff 04/2011
- [24] BILLARI, F. C. (Hrsg.) ; PRSKAWETZ, A. (Hrsg.): *Agent-Based Computational Demography. Using Simulation to Improve Our Understanding of Demographic Behaviour*. Physica-Verlag, 2003
- [25] BIRG, Herwig: *Ursprünge der Demographie in Deutschland. Leben und Werk Johann Peter Süßmilchs (1707-1767)*. Campus Verlag, 1986
- [26] BIRG, Herwig: Betrachtung über die demographischen Aspekte der Ethik und die ethischen Aspekte der Demographie. In: *Bevölkerung. Entwicklung. Umwelt*. Busse Seewald, 1995, S. 257–289
- [27] BIRG, Herwig: *Die Weltbevölkerung. Dynamik und Gefahren*. Verlag C. H. Beck oHG, 2004
- [28] BOGUE, D. J. ; BIEHLER, S.: *Techniques for Making Functional Population Projections*. University of Chicago Community & Family, 1979
- [29] BONGAARTS, John ; FEENEY, Griffith: How Long Do We Live? In: *Population and Development Review* 28 (2002), Nr. 1, S. 13–29
- [30] BOOTH, Heather: Demographic Forecasting: 1980 to 2005 in Review. In: *International Journal of Forecasting* 22 (2006), Nr. 3, S. 547–581
- [31] BOOTH, Heather ; HYNDMAN, Rob J. ; TICKLE, Leonie ; DE JONG, Piet: Lee-Carter mortality forecasting: a multi-country comparison of variants and extensions. In: *Demographic Research* 15 (2006), S. 289–310

- [32] BOOTH, Heather ; MAINDONALD, John ; SMITH, Len: Applying Lee-Carter under conditions of variable mortality decline. In: *Population Studies* 56 (2002), S. 325–336
- [33] BOOTH, Heather ; TICKLE, Leonie: Mortality modelling and forecasting: A review of methods. In: *Annals of Actuarial Science* 3 (2008), Nr. 1-2, S. 3–43
- [34] BOSERUP, Ester: *Population and Technical Change. A Study of Long-term Trends*. University of Chicago Press, 1981
- [35] BOWLEY, A. L.: Births and Population of Great Britain. In: *The Journal of the Royal Economic Society* 34 (1924), S. 188–192
- [36] BOX, G. E. P. ; JENKINS, G. M.: *Time series analysis: Forecasting and Control*. Holden-Day, 1976
- [37] BRASS, W.: Perspectives in Population Prediction: Illustrated by the Statistics of England and Wales. In: *Journal of the Royal Statistical Society Series A (General)* 137 (1974), Nr. 4, S. 532–583
- [38] BRASS, William: On the Scale of Mortality. In: *Biological aspects of demography*, Taylor & Francis LTD, 1971
- [39] BRATLEY, Paul ; FOX, Bennett L. ; SCHRAGE, Linus E.: *A Guide to Simulation. Second Edition*. Springer-Verlag, 1983
- [40] BRETZ, Manfred: Bevölkerungsvorausberechnungen: Statistische Grundlagen und Probleme. In: *Wirtschaft und Statistik* 4 (1986), S. 233–260
- [41] BUNDESAMT FÜR MIGRATION UND FLÜCHTLINGE: *Migrationsbericht des Bundesamtes für Migration und Flüchtlinge im Auftrag der Bundesregierung. Migrationsbericht 2008*. Bundesministerium des Innern, 2010
- [42] BUNDESMINISTERIUM DER JUSTIZ: *Gesetz über die Statistik der Bevölkerungsbewegung und die Fortschreibung des Bevölkerungsstandes (Bevölkerungstatistikgesetz - BevStatG)*. <http://bundesrecht.juris.de/bundesrecht/bevstatg/gesamt.pdf>. Version: 2008. – Zugriff 04/2011
- [43] BUNDESMINISTERIUM FÜR FAMILIE, FRAUEN, SENIOREN UND JUGEND: *Gesetz zum qualitätsorientierten und bedarfsgerechten Ausbau der Tagesbetreuung für Kinder*

- und zur Weiterentwicklung der Kinder- und Jugendhilfe (Tagesbetreuungsausbaugesetz - TAG). <http://www.bmfsfj.de/RedaktionBMFSFJ/Broschuerenstelle/Pdf-Anlagen/Tagesbetreuungsausbaugesetz-TAG,property=pdf,bereich=bmfsfj,sprache=de,rwb=true.pdf>. Version: 2004. – Zugriff 04/2011
- [44] BUNDESMINISTERIUM FÜR FAMILIE, SENIOREN, FRAUEN UND JUGEND: *Familienreport 2009*. Druck Vogt GmbH, 2009
- [45] BÄHR, Jürgen: *Bevölkerungsgeographie*. UTB, 4. Auflage, 2004
- [46] CANNAN, Edwin: The Probability of a Cessation of the Growth of Population in England and Wales during the next Century. In: *The Economic Journal* 5 (1895), Nr. 20, S. 505–515
- [47] CARLIN, Bradley P. ; LOUIS, Thomas A.: *Bayesian Methods for Data Analysis*. Chapman & Hall/CRC, 2008
- [48] CARLSON, Elwood D.: The impact of international migration upon the timing of marriage and childbearing. In: *Demography* 22 (1985), Nr. 1, S. 61–72
- [49] CARNES, Bruce A. ; OLSHANSKY, S. J.: A Realist View of Aging, Mortality, and Future Longevity. In: *Population and Development Review* 33 (2007), Nr. 2, S. 367–381
- [50] CARTER, Lawrence R. ; LEE, Ronald D.: Forecasting demographic components. Modeling and forecasting US sex differentials in mortality. In: *International Journal of Forecasting* 8 (1992), S. 393–411
- [51] CASELLI, Graziella: Future Longevity among the Elderly. In: CASELLI, Graziella (Hrsg.) ; LOPEZ, Alan D. (Hrsg.): *Health and Mortality Among Elderly Populations*. Clarendon Press, 1996
- [52] CASELLI, Graziella ; VALLIN, Jacques ; MARSILI, Marco: How Useful Are the Causes of Death When Extrapolating Mortality Trends. An Update. In: BENGTSSON, Tommy (Hrsg.) ; CHRISTENSEN, Kaare (Hrsg.): *Perspectives on Mortality Forecasting IV. Causes of Death*, Swedish Social Insurance Agency, 2006, 9-36
- [53] CASWELL, Hal: *Matrix Population Models*. Sinauer Associates, Inc., 2001
- [54] CERN - EUROPEAN ORGANIZATION FOR NUCLEAR RESEARCH: *Colt*. <http://acs.lbl.gov/software/colt>. – Zugriff 10/2010

- [55] CERONE, Pietro: On Stable Population Theory with Immigration. In: *Demography* 24 (1987), S. 431–438
- [56] CHANDOLA, T. ; COLEMAN, D. A. ; HIORNS, R. W.: Recent European fertility patterns: Fitting curves to 'distorted' distributions. In: *Population Studies* 53 (1999), Nr. 3, S. 317–329
- [57] CLAYTON, D. ; SCHIFFLERS, E.: Models for temporal variation in cancer rates. I: Age-period and age-cohort models. In: *Statistics in Medicine* 6 (1987), S. 449–467
- [58] CLAYTON, D. ; SCHIFFLERS, E.: Models for temporal variation in cancer rates. II: Age-Period-Cohort Models. In: *Statistics in Medicine* 6 (1987), S. 469–481
- [59] COALE, Ansley J. ; DEMENY, Paul: *Regional Model Life Tables and Stable Populations*. Princeton University Press, 1966
- [60] COALE, Ansley J. ; DEMENY, Paul ; VAUGHN, Barbara: *Regional Model Life Tables and Stable Population*. New York: Academic Press, 1983
- [61] COALE, Ansley J. ; TRUSSELL, T. J.: Model Fertility Schedules: Variations in The Age Structure of Childbearing in Human Populations. In: *Population Index* 40 (1974), Nr. 2, S. 185–258
- [62] COHEN, Joel E.: Population Growth and Earth's Human Carrying Capacity. In: *Science* 269 (1995), Juli, Nr. 5222, S. 341–346
- [63] COHEN, Joel E.: *How Many People Can the Earth Support?* W. W. Norton & Company, Inc., 1996
- [64] CRIMMINS, Eileen M.: The Changing Pattern of American Mortality Decline, 1940-77, and Its Implications for the Future. In: *Population and Development Review* 7 (1981), Nr. 2, S. 229–254
- [65] CRIMMINS, Eileen M.: Trends in the Health of the Elderly. In: *Annual Review of Public Health* 25 (2004), S. 79–98
- [66] CRUIJSEN, Harri ; KEILMAN, Nico: A comparative analysis of the forecasting process. In: KEILMAN, Nico (Hrsg.) ; CRUIJSEN, Harri (Hrsg.): *National Population Forecasting in Industrialized Countries*. Swets & Zeitlinger B. V., 1992, S. 3–25

- [67] DE BEER, Joop: Methods of fertility projections and forecasts. In: KEILMAN, Nico (Hrsg.) ; CRUIJSEN, Harri (Hrsg.): *National Population Forecasting in Industrialized Countries*. Swets & Zeitlinger B. V., 1992, S. 27–48
- [68] DE BEER, Joop: Forecasting International Migration: Time Series Projections vs Argument-Based Forecasts. In: RAYMER, James (Hrsg.) ; WILLEKENS, Frans (Hrsg.): *International Migration in Europe. Data, Models and Estimates*. John Wiley & Sons Ltd., 2008, S. 283–306
- [69] DE BRUIJN, Bart J.: Fertility: Theories, Frameworks, Models, Concepts. In: CASELLI, Graziella (Hrsg.) ; VALLIN, Jacques (Hrsg.) ; WUNSCH, Guillaume (Hrsg.): *Demography: Analysis and Synthesis* Bd. 1. Elsevier Inc., 2005, S. 549–569
- [70] DE BRUIJN, Bart J.: *Foundations of demographic theory. Choice, process, context*. Amsterdam: NethurD-Thela Thesis, 1999
- [71] DE GANS, Henk A.: *Population Forecasting 1895-1945. The Transition to Modernity*. Kluwer Academic Publishers, 1999
- [72] DENEFFE, P. J.: *Die Berechnungen über die künftige deutsche Bevölkerungsentwicklung*. Hans Buske Verlag, 1938
- [73] DIEKMANN, Andreas: *Empirische Sozialforschung. Grundlagen, Methoden, Anwendungen*. Rowohlt Taschenbuch Verlag GmbH, 2001
- [74] DINKEL, Reiner H. ; LEBOK, Uwe H.: The Fertility of Migrants Before and After Crossing the Border: The Ethnic German Population from the Former Soviet Union as a Case Study. In: *International Migration* 35 (1997), Nr. 2, S. 253–268
- [75] DORN, Harald F.: Pitfalls in Population Forecasts and Projections. In: *Journal of the American Statistical Association* 45 (1950), Nr. 251, S. 311–334
- [76] DUBLIN, Louis I. ; LOTKA, Alfred J.: Trends in Longevity. In: *Annals of the American Academy of Political and Social Science* 237 (1945), S. 123–133
- [77] DÜVELL, Franck: *Europäische und internationale Migration. Einführung in historische, soziologische und politische Analysen*. Lit Verlag, 2006
- [78] EASTERLIN, Richard A.: Relative economic status and the American fertility swing. In: SHELDON, Eleanor B. (Hrsg.): *Family Economic Behavior: Problems and Prospects*. Lippincott, 1973



- [79] EDMONSTON, Barry ; PASSEL, Jeffrey S.: Immigration and immigrant generations in population projections. In: *International Journal of Forecasting* 8 (1992), S. 459–476
- [80] ENSTROM, James E. ; BRESLOW, Lester: Lifestyle and reduced mortality among active California Mormons, 1980-2004. In: *Preventive Medicine* (2007), S. 1–4
- [81] ERMISCH, John: Explanatory models for fertility projections and forecasts. In: KEILMAN, Nico (Hrsg.) ; CRUIJSEN, Harri (Hrsg.): *National Population Forecasting in Industrialized Countries*. Swets & Zeitlinger B. V., 1992, S. 201–222
- [82] ESPENSHADE, Thomas J. ; BOUVIER, Leon F. ; ARTHUR, W. B.: Immigration and the Stable Population Model. In: *Demography* 19 (1982), S. 125–133
- [83] EUROSTAT FROM STATISTICS NETHERLANDS (CBS): Demographic statistics: Definitions and methods of collection in 31 European Countries. In: *Working Papers and Studies, Population and social conditions* 3 (2003), Nr. 25. [ec.europa.eu/eurostat/ramon/statmanuals/files/KS-CC-03-005-EN.pdf](http://ec.europa.eu/eurostat/ramon/statmanuals/files/KS-CC-03-005-EN.pdf). – Zugriff 04/2011
- [84] EWALD, Roland ; RÖSSEL, Johannes ; HIMMELSPACH, Jan ; UHRMACHER, Adelinde M.: A plug-in - based architecture for random number generation in simulation systems. In: MASON, S. J. (Hrsg.) ; HILL, R. R. (Hrsg.) ; MOENCH, L. (Hrsg.) ; ROSE, O. (Hrsg.): *Proceedings of the Winter Simulation Conference*, IEEE Computer Society, 2008, 836–844
- [85] FAHLBECK, P. E.: La décadence et la chute des peuples. In: *Bulletin de l'Institut International de Statistique* XV, Annexe 12 (1905), Nr. 2, S. 367–389
- [86] FEYNMAN, Richard P.: *The Pleasure of finding Things out. Doubt and uncertainty*. Interview BBC, Horizon, 1981. <http://www.bbc.co.uk/sn/tvradio/programmes/horizon/broadband/archive/feynman/index.shtml>. – Zugriff 04/2011
- [87] FILDES, Robert: On error measures: A response to the commentators - the best error measure? In: *International Journal of Forecasting* 8 (1992), Nr. 1, S. 109–111
- [88] FU, Michael C.: Optimization for Simulation: Theory vs. Practice. In: *INFORMS Journal on Computing* 14 (2002), Nr. 3, S. 192–215
- [89] FUREDİ, Frank: *Population and Development. A Critical Introduction*. Polity Press, 1997

- [90] GAMPE, Jutta: Human mortality beyond age 110. In: MAIER, Heiner (Hrsg.) ; GAMPE, Jutta (Hrsg.) ; BERNARD, Jeune (Hrsg.) ; ROBINE, Jean-Marie (Hrsg.) ; VAUPEL, James W. (Hrsg.): *Supercentenarians*. Springer, 2010
- [91] GAMPE, Jutta ; ZINN, Sabine: Description of the Microsimulation Model. In: *MicMac Work package 2, MPIDR* (2007). <http://nidi.nl/Content/NIDI/output/micmac/micmac-d8b.pdf>. – Zugriff 04/2011
- [92] GAMPE, Jutta ; ZINN, Sabine ; WILLEKENS, Frans ; VAN DEN GAAG, Nicole: Population forecasting via microsimulation: the software design of the MicMac project. In: EUROSTAT (Hrsg.): *Work session on demographic projections: Bucharest, 10-12 October 2007*. Office for Official Publications of the European Communities, 2007, S. 237–241
- [93] GELMAN, Andrew ; CARLIN, John B. ; STERN, Hal S.: *Bayesian Data Analysis*. CRC Press Inc., 2003
- [94] GEORGE, M. V. ; PERREAULT, J.: Methods of external migration projections and forecasts. In: KEILMAN, Nico (Hrsg.) ; CRUIJSEN, Harri (Hrsg.): *National Population Forecasting in Industrialized Countries*. Swets & Zeitlinger B. V., 1992, S. 87–103
- [95] GILBERT, N. ; TROITZSCH, K. G.: *Simulation for the Social Scientist*. Open University Press, 1999
- [96] GIROSI, Federico ; KING, Gary: Time Series Cross-Sectional Analyses with Different Explanatory Variables in Each Cross-Section. In: *The Global Burden of Disease 2000 in Aging Populations, Research Paper No. 10* (2001)
- [97] GIROSI, Federico ; KING, Gary: *Demographic Forecasting*. Princeton University Press, 2008
- [98] GOMPERTZ, Benjamin: On the Nature of the Function of the Law of Human Mortality, and on a New Mode of Determining the Value of Life Contingencies. In: *Philosophical Transactions of the Royal Society of London* 115 (1825), S. 513–583
- [99] GOODMAN, L. A.: An elementary approach to the population projection matrix, to the population reproductive value, and to related topics in the mathematical theory of population growth. In: *Demography* 5 (1968), S. 382–409
- [100] GRASSBERGER, Peter: On correlations in „good“ random number generators. In: *Physics Letters A* 181 (1993), September, Nr. 1, S. 43–46

- [101] GUILLOT, Michel: Tempo effects in mortality: An appraisal. In: *Demographic Research* 14 (2006), S. 1–26
- [102] GUMBEL, E. J.: Die Versuche eines mathematischen Gesetzes der Bevölkerungszunahme. In: *Allgemeines Statistisches Archiv* (1915), S. 616–656
- [103] HAGE, Friedrich: Vorausberechnungen über Bevölkerungsentwicklung. In: *Deutsches Statistisches Zentralblatt* 23 (1931), S. 1–10
- [104] HALLI, Shiva S. ; VANINADHA, K.: *Advanced Techniques of Population Analysis*. Plenum Press, 1992
- [105] HELIGMAN, L. ; POLLARD, J. H.: The age pattern of mortality. In: *Journal of the Institute of Actuaries* 107 (1980), S. 49–80
- [106] HELLEKALEK, P.: Good random number generators are (not so) easy to find. In: *Mathematics and Computers in Simulation* 46 (1998), Nr. 5-6, S. 485–505
- [107] HERVITZ, Hugo M.: Selectivity, Adaptation, or Disruption? A Comparison of Alternative Hypotheses on the Effects of Migration on Fertility: The Case of Brazil. In: *International Migration Review* 19 (1985), Nr. 2, S. 293–317
- [108] HIBERNATE DEVELOPMENT TEAM: *Hibernate*. <http://hibernate.org>. – Zugriff 10/2010
- [109] HIMMELSPACH, Jan: *Konzeption, Realisierung und Verwendung eines allgemeinen Modellierungs-, Simulations- und Experimentiersystems*, Universität Rostock, Diss., November 2007
- [110] HIMMELSPACH, Jan ; UHRMACHER, Adelinde M.: Plug’n simulate. In: *Proceedings of the 40th Annual Simulation Symposium*, IEEE Computer Society, 2007, 137–143
- [111] HOEM, Jan M. ; MADSEN, Dan ; NIELSEN, Jorgen L. ; OHLSEN, Else-Marie ; HANSEN, Hans O. ; RENNERMALM, Bo: Experiments in Modelling Recent Danish Fertility Curves. In: *Demography* 18 (1981), Nr. 2, S. 231–244
- [112] HOUGAARD, Philip: Multi-state Models: A Review. In: *Lifetime Data Analysis* 5 (1999), S. 239–264
- [113] HOUGAARD, Philip: *Analysis of Multivariate Survival Data*. Springer Science+Business Media, Inc., 2000

- [114] HOWE, Neil ; JACKSON, Richard: Projecting immigration. A survey of the current state of practice and theory. In: *Center for Retirement Research, Working Paper 2004-32* 32 (2004)
- [115] HUININK, Johannes ; WAGNER, Michael: Partnerschaft, Ehe und Familie in der DDR. In: HUININK, Johannes (Hrsg.) ; MAYER, Karl-Ulrich (Hrsg.) ; DIEWALD, Martin (Hrsg.): *Kollektiv und Eigensinn. Lebensläufe in der DDR und danach*. Akademie-Verlag, 1995
- [116] HYNDMAN, Rob J. ; BOOTH, Heather: Stochastic population forecast using functional data models for mortality, fertility and migration. In: *International Journal of Forecasting* 24 (2008), Nr. 3, S. 323–342
- [117] HYNDMAN, Rob J. ; KOEHLER, Anne B.: Another Look at Measures of Forecast Accuracy. In: *International Journal of Forecasting* 22 (2006), Nr. 4, S. 679–688
- [118] HYNDMAN, Rob J. ; ULLAH, Md S.: Robust forecasting of mortality and fertility rates: A functional data approach. In: *Computational Statistics and Data Analysis* 51 (2007), Nr. 10, S. 4942–4956
- [119] HÖHN, Charlotte: Amtliche Bevölkerungsvorausschätzungen seit 1925 - eine kurze Geschichte der Politikberatung und des demographischen Klimas. In: HANAU, Klaus (Hrsg.) ; HUJER, Reinhard (Hrsg.) ; NEUBAUER, Werner (Hrsg.): *Wirtschafts- und Sozialstatistik. Empirische Grundlagen politischer Entscheidungen*. Vandenhoeck & Ruprecht, 1986
- [120] JIOBU, R. ; MARSHALL, H.: Minority Status and Family Size: A Comparison of Explanations. In: *Population Studies* 31 (1977), Nr. 3, S. 509–517
- [121] JONG, Piet de ; TICKLE, Leonie: Extending Lee-Carter Mortality Forecasting. In: *Mathematical Population Studies* 13 (2006), Nr. 1, S. 1–18
- [122] KAHN, Joan R.: Immigrant Selectivity and Fertility Adaption in the United States. In: *Social Forces* 67 (1988), Nr. 1, S. 108–128
- [123] KAHN, Joan R.: Immigrant and Native Fertility during the 1980s: Adaption and Expectations for the Future. In: *International Migration Review* 28 (1994), Nr. 3, S. 501–519

- [124] KANNISTO, Väinö ; LAURITSEN, Jens ; THATCHER, A. R. ; VAUPEL, James W.: Reductions in Mortality at Advanced Ages: Several Decades of Evidence from 27 Countries. In: *Population and Development Review* 20 (1994), Nr. 4, S. 793–810
- [125] KEILMAN, Nico: *Uncertainty in national population forecasting*. Swets and Zeitlinger B. V., 1990
- [126] KEILMAN, Nico: Ex-Post Errors in Official Population Forecasts in Industrialized Countries. In: *Journal of Official Statistics* 13 (1997), S. 245–277
- [127] KEILMAN, Nico: How Accurate Are the United Nations World Population Projections? In: *Population and Development Review, Supplement: Frontiers of Population Forecasting* 24 (1998), S. 15–41
- [128] KEILMAN, Nico: Types of Models for Projecting Mortality. In: BENGTSSON, Tommy (Hrsg.) ; KEILMAN, Nico (Hrsg.): *Perspectives on Mortality Forecasting I. Current Practice*, Swedish National Social Insurance Board, 2003, 19-27
- [129] KEILMAN, Nico: European Demographic Forecasts Have Not Become More Accurate Over the Past 25 Years. In: *Population and Development Review* 34 (2008), Nr. 1, S. 137–153
- [130] KEILMAN, Nico ; PHAM, Dinh Q.: Empirical errors and predicted errors in fertility, mortality and migration forecasts in the European Economic Area. In: *Discussion Papers, Statistics Norway, Social and Demographic Research* (2004), August, Nr. 386. <http://www.ssb.no/publikasjoner/DP/pdf/dp386.pdf>. – Zugriff 04/2011
- [131] KEILMAN, Nico ; PHAM, Dinh Q.: Prediction intervals for Lee-Carter-based mortality forecasts. In: *European Population Conference*, 2006
- [132] KEILMAN, Nico ; PHAM, Dinh Q. ; HETLAND, Arve: Why population forecasts should be probabilistic - illustrated by the case of Norway. In: *Demographic Research* 6 (2002), Nr. 15, S. 409–454
- [133] KEILMAN, Nico ; PHAM, Quang: Predictive Intervals for Age-Specific Fertility. In: *European Journal of Population* 16 (2000), S. 41–66
- [134] KEYFITZ, Nathan: The population projection as a matrix operator. In: *Demography* 1 (1964), S. 56–73

- [135] KEYFITZ, Nathan: Reconciliation of population models: Matrix, integral equation and partial fraction. In: *Journal of the Royal Statistical Society* 130 (1967), S. 61–83
- [136] KEYFITZ, Nathan: The limits of population forecasting. In: *Population and Development Review* 7 (1981), Nr. 4, S. 579–593
- [137] KEYFITZ, Nathan ; CASWELL, Hal: *Applied Mathematical Demography*. Springer Science+Business Media Inc., 2005
- [138] KING, Gavin ; BAUER, Christian ; ANDERSEN, Max R. ; BERNARD, Emmanuel ; EBERSOLE, Steve: *Hibernate Reference Documentation 3.5.6-Final*, 2010. [http://docs.jboss.org/hibernate/core/3.5/reference/en-US/pdf/hibernate\\_reference.pdf](http://docs.jboss.org/hibernate/core/3.5/reference/en-US/pdf/hibernate_reference.pdf). – Zugriff 10/2010
- [139] KNIBBS, George H.: The Laws of Growth of a Population. In: *Journal of the American Statistical Association* 21 (1926), Nr. 156, S. 381–398
- [140] KNUDSEN, Christin ; MCNOWN, Robert ; ROGERS, Andrei: Forecasting Fertility: An Application of Time Series Methods to Parameterized Model Schedules. In: *Social Science Research* 22 (1993), S. 1–23
- [141] KNUTH, Donald E.: *Art of Computer Programming, Volume 2: Seminumerical Algorithms*. 3. Addison-Wesley Professional, 1998
- [142] KOHLS, Martin: Healthy-Migrant-Effect, Erfassungsfehler und andere Schwierigkeiten bei der Analyse der Mortalität von Migranten. In: *Bundesamt für Migration und Flüchtlinge, Working Paper 15* (2008)
- [143] KOHLS, Martin: Leben Migranten wirklich länger? Eine empirische Analyse der Mortalität von Migranten in Deutschland. In: *Bundesamt für Migration und Flüchtlinge, Working Paper 16* (2008)
- [144] KOHLS, Martin: Mortality of Immigrants in Germany. In: *European Population Conference*, 2008
- [145] KONO, S.: Functional population projections. Editor’s introduction. In: BOGUE, Donald J. (Hrsg.) ; ARRIAGA, Eduardo E. (Hrsg.) ; ANDERTON, Douglas L. (Hrsg.) ; RUMSEY, George W. (Hrsg.): *Readings in population research methodology*. Social Development Center, 1993

- [146] KREYENFELD, Michaela ; KONIETZKA, Dirk: Angleichung oder Verfestigung von Differenzen? Geburtenentwicklung und Familienformen in Ost- und Westdeutschland. In: *MPIDR Working Paper WP 2004-025* (2004). <http://www.demogr.mpg.de/papers/working/wp-2004-025.pdf>. – Zugriff 04/2011
- [147] LAND, Kenneth C.: Methods for National Population Forecasts: A Review. In: *Journal of the American Statistical Association* 81 (1986), Nr. 396, S. 888–901
- [148] LANZIERI, Giampaolo: Long-term population projections at national level. In: *Statistics in Focus* 3 (2006)
- [149] LEDERER, Harald W.: *Indikatoren der Migration*. efms, 2004
- [150] LEE, Everett S.: A Theory of Migration. In: *Demography* 3 (1966), Nr. 1, S. 47–57
- [151] LEE, Ronald: The Lee-Carter Method for Forecasting Mortality, with various Extensions and Applications. In: *North American Actuarial Journal* 4 (2000), Nr. 1, S. 80–93
- [152] LEE, Ronald: Quantifying Our Ignorance: Stochastic Forecasts of Population and Public Budgets. In: WAITE, Linda J. (Hrsg.): *Population and Development Review, Supplement: Aging, Health, and Public Policy: Demographic and Economic Perspectives* Bd. 30. Population Council, 2004, S. 153–175
- [153] LEE, Ronald ; TULJAPURKAR, Shripad: Population Forecasting for Fiscal Planning: Issues and Innovations. In: AUERBACH, Alan J. (Hrsg.) ; LEE, Ronald D. (Hrsg.): *Demographic Change and Fiscal Policy*. Cambridge University Press, 2001
- [154] LEE, Ronald D.: Aiming at a Moving Target: Period Fertility and Changing Reproductive Goals. In: *Population Studies* 34 (1980), Nr. 2, S. 205–226
- [155] LEE, Ronald D.: Modeling and forecasting the time series of US fertility: Age distribution, range, and ultimate level. In: *International Journal of Forecasting* 9 (1993), S. 187–202
- [156] LEE, Ronald D.: Probabilistic Approaches to Population Forecasting. In: *Population and Development Review, Supplement: Frontiers of Population Forecasting* 24 (1998), S. 156–190
- [157] LEE, Ronald D. ; CARTER, Lawrence R.: Modeling and Forecasting U.S. Mortality. In: *Journal of the American Statistical Association* 87 (1992), Nr. 419, S. 659–671

- [158] LEE, Ronald D. ; MILLER, Timothy: Evaluating the Performance of the Lee-Carter Method for Forecasting Mortality. In: *Demography* 38 (2001), Nr. 4, S. 537–549
- [159] LEE, Ronald D. ; TULJAPURKAR, Shripad: Stochastic Population Forecasts for the United States: Beyond High, Medium, and Low. In: *Journal of the American Statistical Association* 89 (1994), Nr. 428, S. 1175–1189
- [160] LEE, Ronald D.: Demographic Forecasting and the Easterlin Hypothesis. In: *Population and Development Review* 2 (1976), Nr. 3/4, S. 459–468
- [161] LEISCH, Friedrich: Sweave: Dynamic generation of statistical reports using literate data analysis. In: HÄRDLE, Wolfgang (Hrsg.) ; RÖNZ, Bernd (Hrsg.): *Compstat 2002 - Proceedings in Computational Statistics*, Physica Verlag, Heidelberg, 2002, S. 575–580
- [162] LESLIE, P. H.: On the Use of Matrices in Certain Population Mathematics. In: *Biometrika* 33 (1945), Nr. 3, S. 183–212
- [163] LESLIE, P. H.: Some further notes on the use of matrices in population mathematics. In: *Biometrika* 35 (1948), S. 213–245
- [164] LESLIE, P. H.: The properties of a certain lag type of population growth and the influence of an external random factor on a number of such populations. In: *Physiological Zoology* 32 (1959), S. 151–159
- [165] LESLIE, P. H.: The Intrinsic Rate of Increase and the Overlap of Successive Generations in a Population of Guillemots (*Uria aalge* Pont.). In: *Journal of Animal Ecology* 35 (1966), Nr. 2, S. 291–301
- [166] LEWIS, E. G.: On the generation and growth of a population. In: *Sankhya* 6 (1942), S. 93–96
- [167] LI, Nan ; GERLAND, Patrick: Modifying the Lee-Carter method to project mortality changes up to 2100. In: *PAA Annual Meeting*, 2011
- [168] LONG, John F.: U.S. national population projections methods: A view from four forecasting traditions. In: *Insurance: Mathematics and Economics* 3 (1984), Nr. 4, S. 231–239
- [169] LONG, John F.: Complexity, Accuracy, and Utility of Official Population Projections. In: *Mathematical Population Studies* 5 (1995), Nr. 3, S. 203–216



- [170] LONG, John F. ; McMILLEN, David B.: A survey of Census Bureau population projection methods. In: *Climatic Change* 11 (1987), S. 141–177
- [171] LUTZ, Wolfgang ; GOUJON, Anne: The World's Changing Human Capital Stock: Multi-State Population Projections by Educational Attainment. In: *Population and Development Review* 27 (2001), Nr. 2, S. 323–339
- [172] LUTZ, Wolfgang ; GOUJON, Anne ; DOBLHAMMER-REITER, Gabriele: Demographic Dimensions in Forecasting: Adding Education to Age and Sex. In: *Population and Development Review, Supplement: Frontiers of Population Forecasting* 24 (1998), S. 42–58
- [173] LUTZ, Wolfgang ; SANDERSON, Warren C. ; SCHERBOV, Sergei: Probabilistic population projections based on expert opinion. In: LUTZ, Wolfgang (Hrsg.): *The Future Population of the World. What can we assume today?* Earthscan Ltd, 1996
- [174] LUTZ, Wolfgang ; SANDERSON, Warren C. ; SCHERBOV, Sergei: The end of world population growth. In: *Nature* 412 (2001), August, S. 543–545
- [175] LUTZ, Wolfgang ; SCHERBOV, Sergei: An Expert-Based Framework for Probabilistic National Population Projections: The Example of Austria. In: *European Journal of Population* 14 (1998), Nr. 1, S. 1–17
- [176] LUTZ, Wolfgang ; SCHERBOV, Sergei: Probabilistische Bevölkerungsprognosen für Deutschland. In: *Zeitschrift für Bevölkerungswissenschaft* 23 (1998), Nr. 2, S. 83–109
- [177] LUY, Marc: Mortality tempo-adjustment: An empirical application. In: *Demographic Research* 15 (2006), S. 561–590
- [178] LUY, Marc: Estimating the Migrant Survival Advantage from Orphanhood of Second Generation Migrants. In: *Rostocker Zentrum - Diskussionspapier* (2007), Nr. 17, S. 1–14
- [179] LYNCH, Scott M.: *Introduction to Applied Bayesian Statistics and Estimation for Social Scientists*. Springer Science+Business Media, LLC, 2007
- [180] MALTHUS, Thomas R.: *An Essay on the Principle of Population as it Affects the Future Improvement of Society with Remarks on the Speculations of Mr. Godwin, M. Condorcet and other writers*. London, 1798

- [181] MAMOLO, Marija ; SCHERBOV, Sergei: Population Projections for Forty-Four European Countries: The Ongoing Population Ageing. In: *European Demographic Research Papers* 2 (2009)
- [182] MANTON, Kenneth G. ; PATRICK, Clifford H. ; STALLARD, Eric: Mortality Model Based on Delays in Progression of Chronic Diseases: Alternative to Cause Elimination Model. In: *Public Health Reports* 95 (1980), Nr. 6, S. 580–588
- [183] MANTON, Kenneth G. ; STALLARD, Eric ; TOLLEY, H. D.: Limits to Human Life Expectancy: Evidence, Prospects, and Implications. In: *Population and Development Review* 17 (1991), Nr. 4, S. 603–637
- [184] MASSEY, Douglas S. ; ARANGO, Joaquin ; HUGO, Graeme ; KOUAOUCL, Ali ; PELLEGRINO, Adela ; TAYLOR, J. E.: Theories of International Migration: A Review and Appraisal. In: *Population and Development Review* 19 (1993), Nr. 3, S. 431–466
- [185] MCNOWN, Robert ; ROGERS, Andrei: Forecasting Mortality: A Parameterized Time Series Approach. In: *Demography* 26 (1989), Nr. 4, S. 645–660
- [186] MCNOWN, Robert ; ROGERS, Andrei: Forecasting cause-specific mortality using time series methods. In: *International Journal of Forecasting* 8 (1992), S. 413–432
- [187] MEADOWS, Donella H.: *The Limits to Growth : A Report for the Club of Rome's Project on the Predicament of Mankind*. Universe Publishing, 1972
- [188] MEADOWS, Donella H. ; MEADOWS, Dennis ; RANDERS, Jorgen: *The Limits to Growth: The 30-Year Update*. Chelsea Green Publishing Company, 2004
- [189] MEADOWS, Donella H. ; MEADOWS, Dennis L. ; RANDERS, Jorgen: *Beyond the Limits: Confronting Global Collapse, Envisioning a Sustainable Future*. Chelsea Green Publishing Company, 1992
- [190] MERALI, Zeeya: Computational science: ...Error. In: *Nature* (2010), Oktober, Nr. 467, S. 775–777
- [191] MESLÉ, France: Medical Causes of Death. In: *Demography. Analysis and Synthesis. A Treatise in Population Studies. Volume 2*. Elsevier Inc., 2006
- [192] MILEWSKI, Nadja: First child of immigrant workers and their descendants in West Germany: Interrelation of events, disruption, or adaption? In: *Demographic Research* 17 (2007), Nr. 29, S. 859–896

- [193] MILEWSKI, Nadja: Immigrant fertility in West Germany: Is there a socialisation effect in transitions to second and third births? In: *European Journal of Population* 26 (2010), S. 297–323
- [194] MITRA, S.: Immigration, Below-Replacement Fertility, and Long-Term National Population Trends. In: *Demography* 27 (1990), Nr. 1, S. 121–129
- [195] MITRA, S. ; ROMANIUK, A.: Pearsonian Type I Curve and its Fertility Projection Potentials. In: *Demography* 10 (1973), Nr. 3, S. 351–365
- [196] MITRA, Samarendranath: Generalization of Immigration and the Stable Population Model. In: *Demography* 20 (1983), S. 111–115
- [197] MURRAY, C. J. L. ; FERGUSON, B. D. ; LOPEZ, A. D. ; GUILLOT, M. ; SALOMON, J. A. ; AHMAD, O.: Modified logit life table system: Principles, empirical validation, and application. In: *Population Studies* 57 (2003), Nr. 2, S. 165–182
- [198] MÜLLER, A.: Zur Prognosegüte von Bevölkerungsprognosen. In: *Zeitschrift für Bevölkerungswissenschaft* 13 (1987), Nr. 3, S. 345–378
- [199] NOBLE, Denis: *The Music of Life. Biology beyond the Genome*. Oxford University Press, 2006
- [200] NOTESTEIN, F. W.: Population: the long view. In: SCHULTZ, T. W. (Hrsg.): *Food for the world*, University of Chicago Press, 1945, S. 36–57
- [201] OEPPEN, Jim ; VAUPEL, James W.: Broken Limits to Life Expectancy. In: *Science* 296 (2002), Mai, S. 1029–1031
- [202] OLSHANSKY, S. J.: Simultaneous/multiple Cause-delay (SIMCAD): An Epidemiological Approach to Projecting Mortality. In: *Journal of Gerontology* 42 (1987), Nr. 4, S. 358–365
- [203] OLSHANSKY, S. J. ; CARNES, Bruce A.: Demographic Perspectives on Human Senescence. In: *Population and Development Review* 20 (1994), Nr. 1, S. 57–80
- [204] OLSHANSKY, S. J. ; CARNES, Bruce A.: Prospect for Extended Survival: A Critical Review of the Biological Evidence. In: CASELLI, Graziella (Hrsg.) ; LOPEZ, Alan D. (Hrsg.): *Health and Mortality Among Elderly Populations*. Clarendon Press, 1996
- [205] OLSHANSKY, S. J. ; CARNES, Bruce A.: Ever Since Gompertz. In: *Demography* 34 (1997), Nr. 1, S. 1–15

- [206] OLSHANSKY, S. J. ; CARNES, Bruce A. ; CASSEL, Christine: In Search of Methuselah: Estimating the Upper Limits to Human Longevity. In: *Science* 250 (1990), November, Nr. 4981, S. 634–640
- [207] ORACLE CORPORATION: *Java*. <http://www.oracle.com/technetwork/java/javase/downloads/index.html>. – Zugriff 04/2011
- [208] ORACLE CORPORATION: *Java Swing*. <http://java.sun.com/javase/technologies/desktop>. – Zugriff 10/2010
- [209] ORACLE CORPORATION: *MySQL*. <http://mysql.com/>. – Zugriff 10/2010
- [210] ORCUTT, Guy H.: A new type of socio-economic system. In: *Review of Economics and Statistics* 39 (1957), S. 116–123
- [211] PEARL, Raymond ; REED, Lowell J.: On the Rate of Growth of the Population of the United States since 1790 and its Mathematical Representation. In: *Proceedings of the National Academy of Sciences* 6 (1920), Nr. 6, S. 275–288
- [212] PEDROZA, Claudia: A Bayesian forecasting model: predicting U.S. male mortality. In: *Biostatistics* 7 (2006), Nr. 4, S. 530–550
- [213] PERISTERA, Paraskevi ; KOSTAKI, Anastasia: Modeling fertility in modern populations. In: *Demographic Research* 16 (2007), S. 141–194
- [214] PETTY, Sir William ; GRAUNT, John ; HULL, Charles H.: *The economic writings of Sir William Petty: together with the Observations upon the bills of mortality more probably by Captain John Graunt*. The University Press, 1899
- [215] PFLAUMER, Peter: Confidence intervals for population projections based on Monte Carlo methods. In: *International Journal of Forecasting* 4 (1988), S. 135–142
- [216] PFLAUMER, Peter: *Methoden der Bevölkerungsvorausschätzung unter besonderer Berücksichtigung der Unsicherheit*. Duncker u. Humblot, 1988 (Volkswirtschaftliche Schriften 377)
- [217] PFLAUMER, Peter: Forecasting US population totals with the Box-Jenkins approach. In: *International Journal of Forecasting* 8 (1992), S. 329–338

- [218] PFLAUMER, Peter: Measuring the Accuracy of Population Projections: An Application of Exploratory Data Analysis. In: *Applied stochastic models and data analysis* 2 (1993), S. 755–766
- [219] PHILIPPOV, Dimiter ; ROGERS, Andrei: Multistate Population Projections. In: ROGERS, Andrei (Hrsg.): *Advances in Multiregional Demography*. International Institute for Applied Systems Analysis, 1981, S. 51–82
- [220] POLLARD, John H.: Projection of age-specific mortality rates. In: *Population Bulletin of the United Nations* 21 (1987), Nr. 22, S. 55–69
- [221] PRESTON, Samuel H. ; HEUVELINE, Patrick ; GUILLOT, Michel: *Demography. Measuring and Modeling Population Processes*. Blackwell Publishers, 2001
- [222] PRITCHETT, H. S.: A Formula for Predicting the Population of the United States. In: *Publications of the American Statistical Association* 2 (1891), Nr. 14, S. 278–286
- [223] R DEVELOPMENT CORE TEAM: *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing, 2008. <http://www.R-project.org>. – Zugriff 04/2011
- [224] RAU, Roland ; SOROKO, Eugeny ; JASILIONIS, Domantas ; VAUPEL, James W.: Continued Reductions in Mortality at Advanced Ages. In: *Population and Development Review* 34 (2008), Nr. 4, S. 747–768
- [225] RAVENSTEIN, E. G.: The Laws of Migration. In: *Journal of the Royal Statistical Society* 48 (1885), Nr. 2, S. 167–227
- [226] RAVENSTEIN, E. G.: The Laws of Migration. In: *Journal of the Royal Statistical Society* 52 (1889), Nr. 2, S. 241–301
- [227] RAYMER, James (Hrsg.) ; WILLEKENS, Frans (Hrsg.): *International Migration in Europe. Data, Models and Estimates*. John Wiley & Sons Ltd., 2008
- [228] RAZUM, Oliver: Migration, Mortalität und der Healthy-Migrant-Effekt. In: RICHTER, Matthias (Hrsg.) ; HURRELMANN, Klaus (Hrsg.): *Gesundheitliche Ungleichheit. Grundlagen, Probleme, Perspektiven*. VS Verlag C. H. Beck oHG, 2006, S. 255–270
- [229] RAZUM, Oliver ; ROHRMANN, Sabine: Der Healthy-migrant-Effekt: Bedeutung von Auswahlprozessen bei der Migration und Late-entry-Bias. In: *Gesundheitswesen* 64 (2002), S. 82–88

- [230] REED, Lowell J.: Population Growth and Forecasts. In: *The Annals of the American Academy of Political and Social Science* 188 (1936), S. 159–166
- [231] REESE, George ; YARGER, Randy J. ; KING, Tim ; WILLIAMS, Hugh E.: *Managing and Using MySQL*. O'Reilly Media, 2002
- [232] RICHEL, Yann: *JMathTools*. <http://jmathtools.berlios.de>. – Zugriff 10/2010
- [233] RINNE, Horst: *Wirtschafts- und Bevölkerungsstatistik*. Oldenbourg R. Verlag GmbH, 1996
- [234] ROGERS, Andrei: The multiregional matrix growth operator and the stable interregional age structure. In: *Demography* 3 (1966), Nr. 2, S. 537–544
- [235] ROGERS, Andrei: *Introduction to Multiregional Mathematical Demography*. John Wiley & Sons, Inc., 1975
- [236] ROGERS, Andrei: *Regional population projection models*. Sage Publications, 1985
- [237] ROGERS, Andrei: Parameterized Multistate Population Dynamics and Projections. In: *Journal of the American Statistical Association* 81 (1986), Nr. 393, S. 48–61
- [238] ROGERS, Andrei: Population forecasting: Do simple models outperform complex models? In: *Mathematical Population Studies* 5 (1995), Nr. 3, S. 187–202
- [239] ROGERS, Andrei ; LITTLE, Jani S.: Parameterizing age patterns of demographic rates with the multiexponential model schedule. In: *Mathematical Population Studies* 4 (1994), Nr. 3, S. 175–195
- [240] ROMANIUK, A.: A Three Parameter Model for Birth Projections. In: *Population Studies* 27 (1973), S. 467–478
- [241] RYDER, Norman B.: The Future of American Fertility. In: *Social Problems* 26 (1979), Nr. 3, S. 359–370
- [242] SABOIA, Joao L. M.: Modeling and Forecasting Populations by Time Series: The Swedish Case. In: *Demography* 11 (1974), Nr. 3, S. 483–492
- [243] SABOIA, Joao Luiz M.: Autoregressive Integrated Moving Average (ARIMA) Models for Birth Forecasting. In: *Journal of the American Statistical Association* 72 (1977), Nr. 358, S. 264–270

- [244] SANDERSON, Warren C. ; SCHERBOV, Sergei ; O'NEILL, Brian C. ; LUTZ, Wolfgang: Conditional Probabilistic Population Forecasting. In: *International Statistical Review* 72 (2004), Nr. 2, S. 157–166
- [245] SCHELHASE, Torsten ; RÜBENACH, Stefan P.: Die Todesursachenstatistik - Methodik und Ergebnisse 2004. In: *Wirtschaft und Statistik* 6 (2006), S. 614–629
- [246] SCHERBOV, Sergei ; MAMOLO, Marija ; LUTZ, Wolfgang: Probabilistic Population Projections for the 27 EU Member Countries Based on Eurostat Projections. In: *European Demographic Research Papers, Vienna Institute of Demography* (2008), Nr. 2
- [247] SCHMERTMANN, Carl P.: Immigrant's Ages and the Structure of Stationary Populations with Below-Replacement Fertility. In: *Demography* 29 (1992), Nr. 4, S. 595–612
- [248] SCHMERTMANN, Carl P.: A system of model fertility schedules with graphically intuitive parameters. In: *Demographic Research* 9 (2003), Nr. 5, S. 81–110
- [249] SCHMID, Susanne ; KOHLS, Martin: *Fertilität von Migrantinnen in Deutschland*. Herbsttagung des Arbeitskreises Junge Demographie der DGD, Nürnberg, 2008. [http://www.young-demography.org/docs/0810\\_Nuremberg\\_AutumnMeeting/S\\_Schmid\\_Nuremberg\\_20081009.pdf](http://www.young-demography.org/docs/0810_Nuremberg_AutumnMeeting/S_Schmid_Nuremberg_20081009.pdf). – Zugriff 04/2011
- [250] SCHMID, Susanne ; KOHLS, Martin: *Zwischen Aufnahme- und Herkunftsland - Die Fertilität von Migrantinnen in Deutschland*. Fachkonferenz Demographie von Migranten in Deutschland, Nürnberg, 2010. [http://weltbevoelkerung.org/PDFs/Schmid\\_Kohls\\_100510.pdf](http://weltbevoelkerung.org/PDFs/Schmid_Kohls_100510.pdf). – Zugriff 04/2011
- [251] SCHMITT, Christian: *The Effects of Labour Market Participation on Fertility Decisions: Gender Differences in Cross-National Perspective*, Universität Bielefeld, Diss., 2009. <http://nbn-resolving.de/urn/resolver.pl?urn=urn:nbn:de:hbz:361-14772>. – Zugriff 04/2011
- [252] SCHWARZ, Karl: Vorausschätzung der Bevölkerung des Bundesgebietes bis zum Jahr 2000. In: *Wirtschaft und Statistik* 12 (1963), S. 729–735
- [253] SIMON, Julian L.: *The Ultimate Resource 2*. Princeton University Press, 1998
- [254] SMITH, Stanley K.: Further thoughts on simplicity and complexity in population projection models. In: *International Journal of Forecasting* 13 (1997), S. 557–565

- [255] SMITH, Stanley K. ; SINCICH, Terry: Forecasting state and household populations. Evaluating the forecast accuracy and bias of alternative population projections for states. In: *International Journal of Forecasting* 8 (1992), S. 495–508
- [256] SMITH, Stanley K. ; TAYMAN, Jeff ; SWANSON, David A.: *State and Local Population Projections. Methodology and Analysis*. Kluwer Academic/Plenum Publishers, 2001
- [257] SOBOTKA, Tomas: The rising importance of migrants for childbearing in Europe. In: *Demographic Research* 19 (2008), S. 225–248
- [258] SORMUNEN, J. ; ALHO, Juha: *Technical Report of PEP. Manuscript*, 1999. [http://joyx.joensuu.fi/~ek/pep/userpep.htm#Ch\\_R](http://joyx.joensuu.fi/~ek/pep/userpep.htm#Ch_R). – Zugriff 04/2011
- [259] SPIELAUER, Martin: Microsimulation Modeling of Population, Economic Growth, and Social Security Systems. In: *Interim Report, IR-01-026/Juli, International Institute for Applied Systems Analysis* (2001)
- [260] SPIELAUER, Martin: Dynamic Microsimulation of Health Care Demand, Health Care Finance and the Economic Impact of Health Behavior. Part I: Background and a Comparison with Cell-Based Models. In: *Interim Report, IR-02-032/April, International Institute for Applied Systems Analysis* (2002)
- [261] STATISTICS NETHERLANDS: Changing Population of Europe: Uncertain Future. In: *Final Report* (2005). [http://www.stat.fi/tup/euue/upe\\_final\\_report.pdf](http://www.stat.fi/tup/euue/upe_final_report.pdf). – Zugriff 04/2011
- [262] STATISTISCHES BUNDESAMT: *Vorausschätzung der Bevölkerung für die Jahre 1964 bis 2000*. Fachserie A. Bevölkerung und Kultur. Reihe 1. Bevölkerungsstand und -entwicklung. Sonderbeitrag. Verlag: W. Kohlhammer GmbH, 1963
- [263] STATISTISCHES BUNDESAMT: *Vorausschätzung der Bevölkerung für die Jahre 1966 bis 2000*. Fachserie A. Bevölkerung und Kultur. Reihe 1. Bevölkerungsstand und -entwicklung. Sonderbeitrag. Verlag: W. Kohlhammer GmbH, 1967
- [264] STATISTISCHES BUNDESAMT: *11. koordinierte Bevölkerungsvorausberechnung. Annahmen und Ergebnisse*. 2006 [http://www.destatis.de/jetspeed/portal/cms/Sites/destatis/Internet/DE/Presse/pk/2006/Bevoelkerungsentwicklung/Annahmen\\_und\\_\\_Ergebnisse,property=file.pdf](http://www.destatis.de/jetspeed/portal/cms/Sites/destatis/Internet/DE/Presse/pk/2006/Bevoelkerungsentwicklung/Annahmen_und__Ergebnisse,property=file.pdf). – Zugriff 04/2011



- [265] STATISTISCHES BUNDESAMT: *Bevölkerung Deutschlands bis 2050. 11. koordinierte Bevölkerungsvorausberechnung.* 2006 <http://www.destatis.de/jetspeed/portal/cms/Sites/destatis/Internet/DE/Presse/pk/2006/Bevoelkerungsentwicklung/bevoelkerungsprojektion2050,property=file.pdf>. – Zugriff 04/2011
- [266] STATISTISCHES BUNDESAMT: *Bevölkerung Deutschlands bis 2060. 12. koordinierte Bevölkerungsvorausberechnung.* 2009 [http://www.destatis.de/jetspeed/portal/cms/Sites/destatis/Internet/DE/Presse/pk/2009/Bevoelkerung/pressebroschuere\\_\\_bevoelkerungsentwicklung2009,property=file.pdf](http://www.destatis.de/jetspeed/portal/cms/Sites/destatis/Internet/DE/Presse/pk/2009/Bevoelkerung/pressebroschuere__bevoelkerungsentwicklung2009,property=file.pdf). – Zugriff 04/2011
- [267] STATISTISCHES BUNDESAMT: *Bevölkerung Deutschlands bis 2060. Ergebnisse der 12. koordinierten Bevölkerungsvorausberechnung.* [http://www.destatis.de/jetspeed/portal/cms/Sites/destatis/Internet/DE/Content/Publikationen/Fachveroeffentlichungen/Bevoelkerung/VorausberechnungBevoelkerung/BevoelkerungDeutschland2060\\_\\_5124202099005,property=file.xls](http://www.destatis.de/jetspeed/portal/cms/Sites/destatis/Internet/DE/Content/Publikationen/Fachveroeffentlichungen/Bevoelkerung/VorausberechnungBevoelkerung/BevoelkerungDeutschland2060__5124202099005,property=file.xls). Version: 2009. – Zugriff 04/2011
- [268] STATISTISCHES BUNDESAMT: *Bevölkerung und Erwerbstätigkeit. Wanderungen. Fachserie 1 Reihe 1.2.* <http://www.destatis.de/jetspeed/portal/cms/Sites/destatis/Internet/DE/Content/Publikationen/Fachveroeffentlichungen/Bevoelkerung/Wanderungen/Wanderungen2010120087005.psm1>. Version: 2010. – Zugriff 04/2011
- [269] STATISTISCHES BUNDESAMT: *Datenhandbuch zum Mikrozensus. Scientific Use File 2008*, November 2010. [http://www.forschungsdatenzentrum.de/bestand/mikrozensus/suf/2008/fdz\\_mz\\_suf\\_2008\\_schluesselverzeichnis.pdf](http://www.forschungsdatenzentrum.de/bestand/mikrozensus/suf/2008/fdz_mz_suf_2008_schluesselverzeichnis.pdf). – Zugriff 04/2011
- [270] STATISTISCHES REICHSAMT: Neue Beiträge zum deutschen Bevölkerungsproblem. Die deutschen Familien nach der Volkszählung 1933. Die Lebensbilanz des deutschen Volkes im Jahre 1933. Die biologische Methode der Bevölkerungsvorausberechnung. In: *Sonderhefte zu Wirtschaft und Statistik* (1935), Nr. 15
- [271] STEINBERG, J. ; DOBLHAMMER-REITER, G.: Demografische Bevölkerungsprognosen. Theoretische Grundlagen, Annahmen und Vorhersagesicherheit. In: *Bundesgesundheitsblatt* 53 (2010), S. 393–403

- [272] STOTO, Michael A.: The Accuracy of Population Projections. In: *Journal of the American Statistical Association* 78 (1983), Nr. 381, S. 13–20
- [273] SWANSON, David A. (Hrsg.) ; SIEGEL, Jacob S. (Hrsg.): *The Methods and Materials of Demography*. Academic Press, 2004
- [274] SÜSSMILCH, Johann P.: *Die göttliche Ordnung in den Veränderungen des menschlichen Geschlechts, aus der Geburt, Tod und Fortpflanzung desselben erwiesen*. Berlin, 1741
- [275] TABEAU, Ewa ; EKAMPER, Peter ; HUISMAN, Corina ; BOSCH, Alinda: Improving Overall Mortality Forecasts by Analysing Cause-of-Death, Period and Cohort Effects in Trends. In: *European Journal of Population* 15 (1999), S. 153–183
- [276] TABEAU, Ewa ; VAN DEN BERG JETHS, Anneke ; HEATHCOTE, Christopher ; TABEAU, Ewa (Hrsg.) ; VAN DEN BERG JETHS, Anneke (Hrsg.) ; HEATHCOTE, Christopher (Hrsg.): *Forecasting Mortality in Developed Countries: Insights from a Statistical, Demographic and Epidemiological Perspective*. Kluwer Academic Publishers, Dordrecht, 2001
- [277] THATCHER, A. R.: The Long-Term pattern of adult mortality and the highest attained age. In: *Journal of the Royal Statistical Society. Series A (Statistics in Society)* 162 (1999), Nr. 1, S. 5–43
- [278] THIELE, T. N.: On a Mathematical Formula to express the Rate of Mortality throughout the whole of Life. In: *Journal of the Institute of Actuaries and Assurance Magazine* 16 (1872), S. 313–329
- [279] THOMPSON, Patrick A. ; BELL, William R. ; LONG, John F. ; MILLER, Robert B.: Multivariate Time Series Projections of Parameterized Age-Specific Fertility Rates. In: *Journal of the American Statistical Association* 84 (1989), Nr. 407, S. 689–699
- [280] THOMPSON, Warren S.: Recent Trends in World Population. In: *American Journal of Sociology* (1929), Nr. 34, S. 959–975
- [281] THOMPSON, Warren S. ; WHELPTON, P. K.: Estimates of Future Population of the United States, 1940-2000. In: *National Research Planning Board* (1943). [http://www.archive.org/stream/estimatesoffutur1943thomrich/estimatesoffutur1943thomrich\\_djvu.txt](http://www.archive.org/stream/estimatesoffutur1943thomrich/estimatesoffutur1943thomrich_djvu.txt). – Zugriff 04/2011

- [282] TULJAPURKAR, Shripad ; LEE, Ronald D. ; LI, Qi: Random scenario forecasts versus stochastic forecasts. In: *International Statistical Review* 72 (2004), Nr. 2, S. 185–199
- [283] TULJAPURKAR, Shripad ; LI, Nan ; BOE, Carl: A universal pattern of mortality decline in the G7 countries. In: *Nature* 405 (2000), März, S. 789–792
- [284] UNITED NATIONS: *Model Life Tables for Developing Countries*. New York: United Nations, 1982 (Population Studies no. 77)
- [285] VALLIN, Jacques ; BERLINGUER, Giovanni: From Endogenous Mortality to the Maximum Human Life Span. In: *Demography. Analysis and Synthesis. A Treatise in Population Studies. Volume 2*. Elsevier Inc., 2006, S. 95–115
- [286] VALLIN, Jacques ; MESLÉ, France: The Segmented Trend Line of Highest Life Expectancies. In: *Population and Development Review* 35 (2009), Nr. 1, S. 159–187
- [287] VAN DE KAA, Dirk J.: Europe’s second demographic transition. In: *Population Bulletin* 42 (1987), Nr. 1, S. 1–59
- [288] VAN DE KAA, Dirk J.: Verankerte Geschichten: Ein halbes Jahrhundert Forschung über die Determinanten der Fertilität - Die Geschichte und Ergebnisse. In: *Zeitschrift für Bevölkerungswissenschaft* 1 (1997), S. 3–57
- [289] VAN DER ERF, Rob: International comparison of external migration assumptions. In: KEILMAN, Nico (Hrsg.) ; CRUIJSEN, Harri (Hrsg.): *National Population Forecasting in Industrialized Countries*. Swets & Zeitlinger B. V., 1992, S. 105–126
- [290] VAN IMHOFF , Evert ; KEILMAN, Nico: *LIPRO 2.0: An Application of a Dynamic Demographic Projection Model to Household Structure in the Netherlands*. NIDI/CBGS Publications nr. 23, Swets & Zeitlinger, 1991. – Zugriff 04/2011
- [291] VAN IMHOFF, Evert: The exponential multidimensional demographic projection model. In: *Mathematical Population Studies* 2 (1990), Nr. 3, S. 171–182
- [292] VAN IMHOFF, Evert: A general characterization of consistency algorithms in multidimensional demographic projection models. In: *Population Studies* 46 (1992), S. 159–169
- [293] VAN IMHOFF, Evert ; POST, Wendy: Microsimulation methods for population projections. In: *Population: An English Selection* 10 (1998), S. 97–138

- [294] VAUPEL, J. W.: How Change in Age-specific Mortality Affects Life Expectancy. In: *Population Studies* 40 (1986), S. 147–157
- [295] VAUPEL, James ; MANTON, Kenneth ; STALLARD, Eric: The impact of heterogeneity in individual frailty on the dynamics of mortality. In: *Demography* 16 (1979), Nr. 3, S. 439–454
- [296] VAUPEL, James W.: The Remarkable Improvements in Survival at Older Ages. In: *Philosophical Transactions: Biological Sciences* 352 (1997), Nr. 1363, S. 1799–1804
- [297] VAUPEL, James W.: Demographic Analysis of Aging and Longevity. In: *The American Economic Review* 88 (1998), Nr. 2, S. 242–247
- [298] VAUPEL, James W. ; CAREY, James R. ; CHRISTENSEN, Kaare: It's Never Too Late. In: *Science* 301 (2003), September, S. 1679–1681
- [299] VAUPEL, James W. ; CAREY, James R. ; CHRISTENSEN, Kaare ; JOHNSON, Thomas E. ; YASHIN, Anatoli I. ; HOLM, Niels V. ; IACHINE, Ivan A. ; KANNISTO, Väinö ; KHAZAEI, Aziz A. ; LIEDO, Pablo ; LONGO, Valter D. ; ZENG, Yi ; MANTON, Kenneth G. ; CURTSINGER, James W.: Biodemographic Trajectories of Longevity. In: *Science* 280 (1998), Mai, Nr. 855, S. 855–860
- [300] VAUPEL, James W. ; GAMPE, Jutta: Supercentenarians and the theory of heterogeneity. In: *IUSSP*, 2009
- [301] VAUPEL, James W. ; RAU, Roland ; JASILIONIS, Domantas: The Remarkable, Accelerating Decline in Mortality at Older Ages and the Prospects for Further Improvement in Life Expectancy. In: *Biochimica Clinica* 30 (2006), S. S4–S5
- [302] VERHULST, P.-F.: Notice sur la loi que la population suit dans son accroissement. In: GARNIER, Jean G. (Hrsg.) ; QUETELET, Adolphe (Hrsg.): *Correspondance Mathématique et Physique* Bd. Tome quatrième, Société Belge de Librairie, 1838, S. 113–121
- [303] WESTERGAARD, Harald: The Horoscope of the Population in the Twentieth Century. In: *Bulletin de l'Institut international de Statistique* 17 (1908), S. 103–117
- [304] WHELPTON, P. K.: Population in the United States, 1925–1975. In: *American Journal of Sociology* 34 (1928), Nr. 2, S. 253–270

- [305] WHELPTON, P. K.: An empirical method of calculating future population. In: *Journal of the American Statistical Association* 31 (1936), Nr. 195, S. 457–473
- [306] WHIPPLE, George C.: *Vital Statistics. An introduction to the science of demography. Second Edition.* John Wiley & Sons, Inc., 1923
- [307] WILLEKENS, Frans J.: Demographic forecasting: state of the art and research needs. In: HAZEU, C. A. (Hrsg.) ; FRINKING, G. A. B. (Hrsg.): *Emerging Issues in Demographic Research*, Elsevier Science Publishers B. V., 1990, S. 9–66
- [308] WILLEKENS, Frans J.: Biographic forecasting: bridging the micro-macro gap in population forecasting. In: *New Zealand Population Review* 31 (2005), Nr. 1, S. 77–124
- [309] WILLEKENS, Frans J.: Description of the micro-simulation model (Continuous-time micro-simulation). In: *MicMac Work package 2, NIDI* (2006). <http://www.nidi.knaw.nl/Content/NIDI/output/micmac/micmac-d8a.pdf>. – Zugriff 04/2011
- [310] WILLEKENS, Frans J.: Description of the multistate projection model (Multistate model for biographic analysis and projection). In: *MicMac Work package 1, NIDI* (2006). <http://www.nidi.nl/Content/NIDI/output/micmac/micmac-d1.pdf>. – Zugriff 04/2011
- [311] WILLIAMS, George C.: Pleiotropy, Natural Selection, and the Evolution of Senescence. In: *Evolution* 11 (1957), Nr. 4, S. 398–411
- [312] WILMOTH, J. R.: Demography of longevity: past, present, and future trends. In: *Experimental Gerontology* 35 (2000), S. 1111–1129
- [313] WILMOTH, J. R. ; DEEGAN, L. J. ; LUNDSTRÖM, H. ; HORIUCHI, S.: Increase of Maximum Life-Span in Sweden, 1861-1999. In: *Science* 289 (2000), September, Nr. 5488, S. 2366–2368
- [314] WILMOTH, John R.: Are mortality projections always more pessimistic when disaggregated by cause of death? In: *Mathematical Population Studies* 5 (1995), Nr. 4, S. 293–319
- [315] WILMOTH, John R.: Mortality Projections for Japan: A Comparison of Four Methods. In: CASELLI, Graziella (Hrsg.) ; LOPEZ, Alan D. (Hrsg.): *Health and Mortality Among Elderly Populations.* Clarendon Press, 1996

- [316] WILMOTH, John R.: In Search of Limits. In: WACHTER, Kenneth W. (Hrsg.) ; FINCH, Caleb E. (Hrsg.): *Between Zeus and the Salmon*. National Academy Press, 1997
- [317] WILMOTH, John R.: How Long Can We Live? A Review Essay. In: *Population and Development Review* 27 (2001), Nr. 4, S. 791–800
- [318] WITT, Cai D.: Vorausschätzung der Bevölkerung für die Jahre 1966 bis 2000. In: *Wirtschaft und Statistik* 11 (1966), S. 677–683
- [319] WUNSCH, Guillaume: Dependence and Independence of Causes of Death. In: *Demography. Analysis and Synthesis. A Treatise in Population Studies. Volume 2*. Elsevier Inc., 2006, S. 57–60
- [320] YOUNG, Christabel M.: Changes in the Demographic Behaviour of Migrants in Australia and the Transition Between Generations. In: *Population Studies* 45 (1991), Nr. 1, S. 67–89
- [321] ZEIGLER, Bernard P. ; KIM, Tag-Gon ; PRAEHOFER, Herbert: *Theory of Modeling and Simulation. Integrating Discrete Event and Continuous Complex Dynamic Systems*. Academic Press, 2000
- [322] ZEMEK, Cameron: *Jeppers*. <http://jeppers.sourceforge.net>. – Zugriff 10/2010
- [323] ZINN, Sabine ; HIMMELSPACH, Jan ; GAMPE, Jutta ; UHRMACHER, Adelinde M.: MIC-CORE: A Tool for Microsimulation. In: ROSSETTI, M. D. (Hrsg.) ; HILL, R. R. (Hrsg.) ; JOHANSSON, B. (Hrsg.) ; DUNKIN, A. (Hrsg.) ; INGALLS, R. G. (Hrsg.): *Proceedings of the Winter Simulation Conference*, IEEE Computer Society, 2009, S. 992–1002

# Abbildungsverzeichnis

2.1.	Grundlegende Terminologie von Bevölkerungsprognosen . . . . .	17
2.2.	Genereller Ablauf eines Prognoseprozesses . . . . .	19
3.1.	Kohorten-Komponenten Methode in Matrixnotation: Aufbau der Projektionsmatrix und der entsprechenden Bevölkerungsvektoren . . . . .	37
3.2.	Ablauf einer deterministischen Bevölkerungsprognose . . . . .	42
3.3.	Ablauf einer probabilistischen Bevölkerungsprognose . . . . .	44
4.1.	Wesentliche Unterscheidungsmerkmale probabilistischer Ansätze zur Bevölkerungsprognose . . . . .	58
5.1.	Altersspezifische Sterbewahrscheinlichkeiten der Männer . . . . .	80
5.2.	Transformation (altersspezifischer) Sterbewahrscheinlichkeiten . . . . .	82
5.3.	Bestimmung der zentrierten Sterberaten, $\ln m_{x,t} - \alpha_x$ , über die Randverteilungen $k_t$ und $\beta_x$ im Lee-Carter Modell . . . . .	86
6.1.	Entwicklung der TFR in den alten und neuen Bundesländern von 1955 bis 2008 . . . . .	112
6.2.	Altersspezifische Fertilitätsraten in den alten und neuen Bundesländern 1990 und 2008 . . . . .	113
7.1.	Entwicklung der Zu- und Fortzüge und des Wanderungssaldos über die Außengrenzen Deutschlands in der Zeit von 1991 bis 2008 . . . . .	129
7.2.	Matrixnotation eines interregionalen Bevölkerungsprognosemodells, Teil 1 .	136
7.3.	Matrixnotation eines interregionalen Bevölkerungsprognosemodells, Teil 2 .	137
8.1.	(Nicht)Modellierbarkeit der Abweichungen (un)ähnlicher Bevölkerungsentwicklungen mit einer linearen Regression . . . . .	161
9.1.	Aspekte des theoretischen Modells, PPPM, und seiner Realisierung P3J . .	167

9.2.	Modellparameter des PPPM bzw. Dateneingabe im P3J . . . . .	168
9.3.	Aufbau des PPPM hinsichtlich zu unterscheidender Subpopulationen . . . .	169
9.4.	Aufbau des PPPM hinsichtlich zu unterscheidender Subpopulationen und Modellparameter . . . . .	177
9.5.	Screenshot der P3J Eingabemaske für einen Annahmenverlauf . . . . .	181
10.1.	Berechnung grundlegender Ergebnisgrößen im PPPM bzw. P3J . . . . .	183
10.2.	Reihenfolge demographischer Ereignisse im PPPM . . . . .	186
10.3.	Zufällige Auswahl eines Annahmenverlaufs im Open Type . . . . .	209
10.4.	Zufällige Auswahl eines Annahmenverlaufs im Limited Type . . . . .	211
10.5.	Realisierung der P3J-Funktionen mit JAMES II . . . . .	217
11.1.	Prognoseergebnisse im PPPM bzw. P3J . . . . .	223
11.2.	Ergebnisverteilung der Gesamtbevölkerung in der Variante II der Beispielprojektion über den Projektionshorizont . . . . .	227
11.3.	Wahrscheinlichkeitsverteilung der Gesamtbevölkerung in der Variante II der Beispielprojektion von 2010 bis 2050 . . . . .	229
11.4.	Bevölkerungspyramide der Gesamtbevölkerung im Jump-off Jahr 2006 und Zieljahr 2050 der Variante II der Beispielprojektion . . . . .	230
11.5.	Kumulierte Wahrscheinlichkeitsverteilung der Gesamtbevölkerung im Zieljahr der Variante II der Beispielprojektion und ausgewählter Miniprojektionen . . . . .	232
11.6.	Gesamtbevölkerung der Variante II der Beispielprojektion und ausgewählter Miniprojektionen über den Projektionshorizont . . . . .	233
11.7.	Levelplot mit den zufällig gezogenen Annahmenverläufen aller 78 Modellparameter für alle aufsteigend sortierten Prognosewerte der Gesamtbevölkerung in der blauen Miniprojektion . . . . .	236
11.8.	Modul für die Generierung eines Ergebnisreports . . . . .	240
12.1.	Bestimmung der Jump-off Werte für die altersspezifischen Sterbewahrscheinlichkeiten der direkten Zuzüge und ihrer ersten Nachkommengeneration . . . . .	244
12.2.	Annahmenverläufe zur Sterblichkeit der Einheimischen . . . . .	246
12.3.	Dunkelblauer, roter und grüner Annahmenverlauf zur Sterblichkeit der einheimischen Männer . . . . .	247



## ABBILDUNGSVERZEICHNIS

---

12.4.	Annahmenverläufe zur Sterblichkeit für die Immigrantinnen . . . . .	250
12.5.	Annahmenverläufe zur Sterblichkeit für die Immigranten . . . . .	252
12.6.	Bestimmung der Jump-off Werte für die altersspezifischen Fertilitätsraten der direkten Zuzüge und ihrer ersten Nachkommengeneration . . . . .	254
12.7.	Annahmenverläufe zur Fertilität für die einheimischen Frauen . . . . .	255
12.8.	Dunkelblauer, blauer, roter und grüner Annahmenverlauf zur Fertilität der einheimischen Frauen . . . . .	257
12.9.	Annahmenverläufe zur Fertilität für die Immigrantinnen . . . . .	258
12.10.	Annahmenverläufe für die Zuzüge nach Geschlecht . . . . .	260
12.11.	Annahmenverläufe für die Fortzüge nach Geschlecht . . . . .	261
12.12.	Alters- und Geschlechtsstruktur der Zu- und Fortzüge im Jahr 2006 . . . .	262
12.13.	Grüne Annahmenverläufe der jährlichen Zu- und Fortzüge für die Frauen .	263
12.14.	Grüner bzw. dunkelblauer Annahmenverlauf der jährlichen Zu- bzw. Fortzüge der Männer . . . . .	264
12.15.	Wanderungssaldo . . . . .	265
12.16.	Annahmenverlauf zum <i>Anteil der Säuglingssterbefälle im ersten Halbjahr</i> .	266
12.17.	Merkmale der vier Varianten der Beispielprojektion . . . . .	268
12.18.	Screenshot: Erstellung einer neuen Projektion im P3J . . . . .	268
12.19.	Screenshot: Grundeinstellungen für eine neue Projektion im P3J . . . . .	269
12.20.	Screenshot: Anordnung der Modellparameter und Annahmenverläufe in einer Baumstruktur im P3J . . . . .	270
12.21.	Screenshot: Eingabe eines neuen Annahmenverlaufs im P3J . . . . .	270
12.22.	Screenshot: Einstellungen für einen neuen Annahmenverlauf im P3J . . . .	271
12.23.	Screenshot: Grundeinstellungen für einen neuen Settyp im P3J . . . . .	272
12.24.	Screenshot: Erstellung eines neuen Sets im P3J . . . . .	273
12.25.	Screenshot: Einstellungen für ein neues Set im P3J . . . . .	273
12.26.	<i>Einfache</i> und <i>repräsentative</i> Annahmenverteilung zur Mortalität der einheimischen Männer . . . . .	277
12.27.	Komplexe <i>repräsentative</i> Annahmenverteilung zur Mortalität der einheimischen Männer . . . . .	278
12.28.	Screenshot: Bestimmung der Anzahl der Prognosedurchläufe und des Berechnungsmodus im P3J . . . . .	279
12.29.	Screenshot: Einstellung der Anzahl der Prognosedurchläufe und des Berechnungsmodus im P3J . . . . .	279

## ABBILDUNGSVERZEICHNIS

---

12.30. Screenshot: Exportieren bzw. aggregieren der Prognoseergebnisse . . . . .	280
12.31. Entwicklung der Gesamtbevölkerung über den Projektionshorizont in den Varianten I bis IV der Beispielpjektion . . . . .	281
12.32. Kumulierte Wahrscheinlichkeitsverteilung der Gesamtbevölkerung im Zieljahr in den Varianten I bis IV der Beispielpjektion . . . . .	282
12.33. Dichteverteilung der Gesamtbevölkerung im Zieljahr in den Varianten I bis IV der Beispielpjektion . . . . .	283
12.34. Entwicklung der Gesamtbevölkerung von 2009 bis 2060 in der 12. koordinierten Bevölkerungsprognose vom Statistischen Bundesamt und in der Variante II der Beispielpjektion . . . . .	284
12.35. Entwicklung einzelner (Sub)Populationen von 2007 bis 2050 in der Variante II der Beispielpjektion . . . . .	285
12.36. Alters- und Geschlechtsstruktur der ersten und zweiten Nachkommengeneration der direkten Zuzüge im Zieljahr 2050 der Variante II der Beispielpjektion . . . . .	286
12.37. Gesamtbevölkerung in der Variante II der Beispielpjektion nach 6.000 quasi-vollständig bzw. simulativ ermittelten Ergebnisverläufen . . . . .	288
13.1. Ablauf einer Bevölkerungsprognose . . . . .	291
B.1. Alter der Personen zur Jahresmitte und zum Jahresende bzw. -anfang im PPPM . . . . .	346
B.2. Bestimmung eines Alters für die Zu- und Fortzüge nach der amtlichen Statistik . . . . .	347
B.3. Festlegung der Zeitpunkte für die Migration im PPPM . . . . .	348
B.4. Festlegung der Zeitpunkte für die Migration und die Mortalität im PPPM .	349
B.5. Verschiebung des Gültigkeitsbereichs der halbjährigen Überlebenswahrscheinlichkeiten im PPPM . . . . .	350
B.6. Tatsächlicher Gültigkeitsbereich der halbjährigen Überlebenswahrscheinlichkeiten im PPPM . . . . .	350
C.1. Stochastische Variation des blauen repräsentativen Annahmenverlaufs zur Sterblichkeit der einheimischen Männer . . . . .	352

# Tabellenverzeichnis

10.1.	Überblick über die Formeln zur Berechnung der Jahresdurchschnittsbestände für die Einheimischen und die Nachkommengenerationen der direkten Zu- und Fortzüge . . . . .	198
10.2.	Überblick über die Formeln zur Berechnung der Jahresdurchschnittsbestände für die direkten Zu- und Fortzüge . . . . .	199
10.3.	Überblick über die Formeln zur Berechnung der Jahresendbestände für die Einheimischen und die Nachkommengenerationen der direkten Zu- und Fortzüge . . . . .	199
10.4.	Überblick über die Formeln zur Berechnung der Jahresendbestände für die direkten Zu- und Fortzüge . . . . .	199
11.1.	Prognostizierte Jahresend- und Jahresdurchschnittsbestände für die einzeln prognostizierten und daraus variabel aggregierten (Sub)Populationen . . . .	225
11.2.	Reihenfolge der 78 Modellparameter im Levelplot (der Abbildung 11.7) für die Gesamtbevölkerung der blauen Miniprojektion . . . . .	235
12.1.	Annahmenverläufe zur Mortalität der einheimischen Frauen und Männer im Zieljahr . . . . .	245
12.2.	Annahmenverläufe zur Mortalität der Immigrantinnen im Zieljahr . . . . .	251
12.3.	Annahmenverläufe zur Mortalität der Immigranten im Zieljahr . . . . .	251
12.4.	Annahmenverläufe zur Fertilität der Einheimischen im Zieljahr . . . . .	254
12.5.	Annahmenverläufe zur Fertilität der Immigrantinnen im Zieljahr . . . . .	259
12.6.	Annahmenverläufe zur Bruttomigration im Zieljahr . . . . .	262
12.7.	Settyp Fertilität in der Variante II der Beispielprojektion . . . . .	274
12.8.	Settyp Mortalität in der Variante II der Beispielprojektion . . . . .	275
12.9.	Settyp Migration in der Variante II der Beispielprojektion . . . . .	276
12.10.	Repräsentative Annahmenverläufe in der Variante III der Beispielprojektion	276

# A. Installation des P3J

An dieser Stelle erfolgt eine kurze Installationsanleitung für das P3J, das auf der CD-ROM (in dieser Arbeit) enthalten ist. Es wird erklärt welche zusätzlichen Komponenten zur Benutzung des P3J benötigt werden und wie diese für ein nahtloses Zusammenspiel mit dem P3J zu konfigurieren sind.

## A.1. Benötigte Software

Um das P3J von der Installations-CD nutzen zu können, müssen zunächst eine Java Runtime Environment (JRE) und ein Datenbankmanagementsystem (wie z. B. MySQL) installiert werden.

### A.1.1. Java

Das P3J ist in Java 1.6 geschrieben, daher wird eine Java-Laufzeitumgebung (JRE) Version 1.6 oder höher zur Ausführung benötigt.<sup>1</sup>

### A.1.2. MySQL

Prinzipiell kann das P3J mit jedem Datenbankmanagementsystem, das von Hibernate unterstützt wird, betrieben werden.<sup>2</sup>

Für die Entwicklung und das Testen des P3J ist jedoch bisher ausschließlich MySQL verwendet worden, sodass eine Installation von MySQL (ab Version 5.0) empfohlen wird.<sup>3</sup> Die Verwendung von MySQL ist bereits voreingestellt und sollte ohne weitere Einstellungen

---

<sup>1</sup>Solche JREs stehen für viele verschiedene Betriebssysteme unter <http://www.oracle.com/technetwork/java/javase/downloads/> zur Verfügung.

<sup>2</sup>Eine Liste mit unterstützten DBMS findet sich beispielsweise bei King et al. ([138], S. 41)

<sup>3</sup>Eine kostenlose Version ist unter <http://www.mysql.de> erhältlich.

funktionieren, abgesehen von den Schritten in Abschnitt A.3.1.

Zur Verwendung eines anderen DBMS sollte es genügen, die Einstellung `dialect` in der Datei `./hibernate.cfg.xml` (von der Installations-CD) von `org.hibernate.dialect.MySQL5Dialect` auf einen anderen Dialekt umzustellen, einen entsprechenden JDBC-Treiber im Java-Klassenpfad zur Verfügung zu stellen, und die Datenbank-URL in der Datei `./config.xml` anzupassen (siehe Abschnitt A.3.1).

### A.1.3. R und L<sup>A</sup>T<sub>E</sub>X

Für die Generierung von Ergebnisreports werden außerdem noch die quelloffene Statistik-Programmierungsumgebung R (<http://www.r-project.org>) sowie das ebenfalls quelloffene Textsatzsystem L<sup>A</sup>T<sub>E</sub>X (<http://www.latex-project.org/>) benötigt.

## A.2. Benötigte Java-Komponenten

Das P3J verwendet zahlreiche Hilfskomponenten, die in einem gesonderten Verzeichnis (`./libs/`) auf der Installations-CD bereitgestellt werden. Diese Bibliotheken sind zwar quell-offen und kostenlos verfügbar, haben jedoch verschiedene Lizenzen. Die entsprechende Lizenz ist jeweils in der Datei `LICENSE.TXT` gespeichert:

- Hibernate, Version 3.5.6  
Verzeichnis: `./libs/hibernate` (siehe <http://www.hibernate.org>)  
Abhängigkeiten zu:
  - Simple Logging Facade for Java (SLF4J), Version 1.5.8  
Verzeichnis: `./libs/slf4j` (siehe <http://www.slf4j.org>)
  - Java Transaction API (JTA), Version 1.1  
Verzeichnis: `./libs/jta` (siehe <http://www.oracle.com/technetwork/java/javasee/tech>)
  - ANother Tool for Language Recognition (ANTLR), Version 2.7.6  
Verzeichnis: `./libs/antlr` (siehe <http://www.antlr.org>)
  - Apache Commons Collections, Version 3.1  
Verzeichnis: `./libs/commons_collections` (siehe <http://commons.apache.org/collections>)

- DOM4J, Version 1.6.1  
Verzeichnis: `./libs/dom4j` (siehe <http://dom4j.sourceforge.net>)
- Java Programming Assistant (Javassist), Version 3.9.0  
Verzeichnis: `./libs/javassist` (siehe <http://jboss.org/javassist>)
- Freemarker, Version 2.3.16  
Verzeichnis: `./libs/freemarker` (siehe <http://freemarker.sourceforge.net/>)
- JGoodies Forms & Looks, Version 1.2.1 (Forms) bzw. 2.3.1 (Looks)  
Verzeichnis: `./libs/jgoodies` (siehe <http://www.jgoodies.com/freeware>)
- Colt, Version 1.2.0  
Verzeichnis: `./libs/colt` (siehe <http://acs.lbl.gov/software/colt>)
- JAVa-based Multipurpose Environment for Simulation (JAMES) II, Version 0.8.3  
Verzeichnis: `./` und `./libs/james`<sup>4</sup> (siehe <http://jamesii.org>)
- MySQL JDBC Connector, Version 5.0.7.  
Verzeichnis: `./libs/mysql_conn` (siehe <http://www.mysql.com/products/connector>)
- Jeppers JGrid, Version 1.1  
Verzeichnis: `./libs/jeppers_jgrid` (siehe <http://jeppers.sourceforge.net>)
- JMathPlot, Version 1.0  
Verzeichnis: `./libs/jmathplot` (siehe <http://sourceforge.net/projects/jmathplot/>)

## A.3. Ausführung

### A.3.1. Vorbereitungen

#### Datenbank-Schema anlegen

Zunächst sollte der MySQL-Datenbankserver gestartet werden (sofern dies nicht automatisch geschieht). Danach ist es nötig, ein Datenbankschema anzulegen, in welchem das P3J die gespeicherten Daten ablegen kann. Dies kann von der Kommandozeile aus oder mit graphischen

---

<sup>4</sup>Die `.jar`-Datei, die den JAMES II-Kern enthält, liegt im Hauptverzeichnis, sodass alle Plug-ins ordnungsgemäß geladen werden können. Die JAMES II-Lizenz befindet sich unter `./libs/james`.

Werkzeugen wie dem MySQL Administrator<sup>5</sup> geschehen. Über die Kommandozeile reichen die folgenden Schritte<sup>6</sup>:

1. Verbindung zum MySQL-Server öffnen:

```
> mysql -h localhost -u mustermann -p
```

2. Nach Eingabe des Passworts für Nutzer 'mustermann'<sup>7</sup> erscheint das MySQL-Kommandoprompt (`mysql>`).

3. Nun muss das Datenbankschema angelegt werden:

```
mysql> create schema neues_pppm_schema;
```

4. Die Existenz des neuen Schemas kann mit

```
show schemas;
```

überprüft werden.

Aus Sicherheitsgründen ist es außerdem empfehlenswert, einen gesonderten Datenbank-Nutzer anzulegen, der lediglich über Lese- und Schreibrechte auf dem neu angelegten Datenbank-Schema verfügt.

### P3J Konfigurationsdatei anpassen

Nachdem die zu benötigende Software installiert und eingestellt ist, können die Dateien von der Installations-CD in ein lokales Verzeichnis gespeichert werden. Unter `./config.xml` sollten nun abschließend die Verbindungsdaten der Datenbank eingestellt werden:

- Eigenschaft **Database URL**: hier muss die URL zum Datenbankschema eingegeben werden, im Falle einer lokalen Installation (auf `localhost`) und einem wie in Abschnitt A.3.1 angelegten Schema 'neues\_pppm\_schema' sollte die entsprechende Zeile auf

```
<string>jdbc:mysql://localhost/neues_pppm_schema</string>
```

geändert werden.

---

<sup>5</sup><http://dev.mysql.com/doc/administrator/en/>

<sup>6</sup>Hier wird von einer lokalen Installation von MySQL ausgegangen. Sollte statt dessen ein Datenbankserver auf einem anderen System verwendet werden, muss `localhost` durch eine entsprechende URL ersetzt werden.

<sup>7</sup>Normalerweise hat der Administrationsaccount die Bezeichnung `root`.

- Eigenschaft `Database User Name`: der Nutzername.
- Eigenschaft `Database Password`: das Passwort des Nutzers. **Achtung: das Passwort wird in dieser Datei im Klartext gespeichert.** Das bedeutet, jeder der Lesezugriff auf die Datei `config.xml` hat kann sich mit dem P3J-Nutzernamen bei der Datenbank anmelden. Die Leserechte für diese Datei sollten dementsprechend eingeschränkt sein (müssen jedoch noch einen Lese-/Schreibzugriff des P3J selbst erlauben).

#### A.3.2. P3J starten

Um das P3J zu starten stehen zwei Startdateien, eine für Windows-Systeme (`./p3j.bat`) und eine für Unix-Systeme (`./p3j.sh`), im Hauptverzeichnis bereit.

### A.4. Technische Dokumentation

Im Verzeichnis `./techdoc/` befinden sich automatisch generierte Zusammenfassungen und Prüfberichte zum Quelltext des Programms. Diese sollen Entwicklern einen schnellen und einfachen Überblick über die Quelldateien ermöglichen. Zum Aufruf einer Übersicht sollte die Datei `./techdoc/index.html` in einem WWW-Browser aufgerufen werden.



## B. Anpassung der Reihenfolge der demographischen Ereignisse im PPPM an die Datenerfassung in Deutschland

An dieser Stelle wird die Abstimmung der Reihenfolge der demographischen Ereignisse im PPPM auf die Datenerfassung in Deutschland genauer beleuchtet.

Im PPPM soll die Reihenfolge der demographischen Ereignisse möglichst realitätsnah modelliert werden und mit den erfassten demographischen Daten der amtlichen Statistik in Deutschland zusammenpassen. Demographische Ereignisse wie Sterbefälle, Geburten, Zuzüge und Fortzüge finden in der Realität über das ganze Jahr verteilt statt.<sup>1</sup> Um diesen eigentlich kontinuierlich auftretenden demographischen Ereignissen mit dem diskreten PPPM<sup>2</sup> gerecht zu werden, findet eine Aufspaltung einiger demographischer Ereignisse in einem Altersintervall statt. Da Geburten über das gesamte Kalenderjahr stattfinden, wird im PPPM die Jahresmitte zur Realisierung der Geburten festgelegt. Daraus ergibt sich, dass sich die zu prognostizierende Bevölkerung im PPPM zur Jahresmitte im exakten Alter  $x$ ,  $x, 0$ , befindet. Folglich befindet sich die zu prognostizierende Bevölkerung zum Jahresende immer im Alter  $x, 5$  (vgl. Abbildung B.1).

Diese Annahme zum Alter der Bevölkerung zur Jahresmitte und zum Jahresende bzw. -anfang korrespondiert mit der relativ groben Zuordnung des Alters bei Migranten in der amtlichen Statistik. In der amtlichen Statistik wird nämlich allen Migranten eines Geburtsjahrgangs das Alter, dass sie alle am Jahresende des jeweiligen Beobachtungsjahres innehaben, zugeordnet. Folglich wird das Alter eines Zu- oder auch eines Fortzugs über die Differenz aus Beobachtungs- und Geburtsjahr berechnet [268]. Tatsächlich variieren

---

<sup>1</sup>Für jedes demographische Ereignis gibt es zwar eine spezifische Saisonfigur, die hier aber aus Gründen der Vereinfachung vernachlässigt werden soll.

<sup>2</sup>Die Projektionsintervalle umfassen in der Regel ein Jahr.

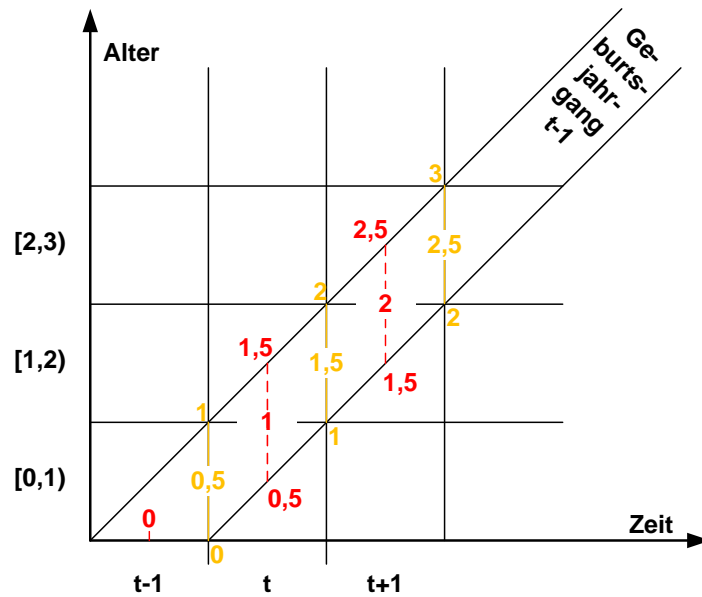


Abbildung B.1.: Alter der Personen zur Jahresmitte und zum Jahresende bzw. -anfang im PPPM: die Jahresdurchschnittsbestände (rot) sind im Durchschnitt immer exakt  $x,0$  Jahre alt, wohingegen die Jahresend- und Jahresanfangsbestände (orange) im Durchschnitt immer  $x,5$  Jahre alt sind.

aber die Alter der Migranten eines Geburtsjahrgangs in einem Beobachtungsjahr über eine Spanne von beinahe zwei Altersjahren. Abbildung B.2 verdeutlicht diesen Sachverhalt. Laut der amtlichen Statistik in Deutschland sind alle Migranten des Geburtsjahrgangs  $t - 1$  im Beobachtungsjahr  $t$  im Altersintervall von 1 bis unter 2 Jahren. In Wirklichkeit sind die Migranten des Geburtsjahrgangs  $t - 1$  im Beobachtungsjahr  $t$  aber zwischen einem Tag und fast 2 Jahren alt, sodass sie nur im Durchschnitt<sup>3</sup> 1 Jahr alt sind.<sup>4</sup>

Um dieser Diskrepanz entgegenzuwirken, findet die Migration im PPPM zu Beginn und zum Ende eines Beobachtungsjahres statt. Die erste Hälfte der Zu- und Fortzüge findet demnach im PPPM zu Beginn des Beobachtungsjahres  $t$  für den Geburtsjahrgang  $t - 1$  statt und die zweite Hälfte findet zum Ende des Beobachtungsjahres  $t$  für den Geburtsjahrgang  $t - 1$  statt (vgl. Abbildung B.3). Damit kann der Migration des Geburtsjahrgangs  $t - 1$ , die über das gesamte Beobachtungsjahr  $t$  verteilt stattfindet, zumindest teilweise entsprochen

<sup>3</sup>bei Annahme einer Gleichverteilung der Migranten nach Alter eines Geburtsjahrgangs im Beobachtungsjahr

<sup>4</sup>Die Personen des Geburtsjahrgangs  $t - 1$  sind zur Jahresmitte des Beobachtungsjahrs  $t$  zwischen 0,5 und 1,5 Jahre alt, sodass sie dann im Durchschnitt 1 Jahr alt sind.

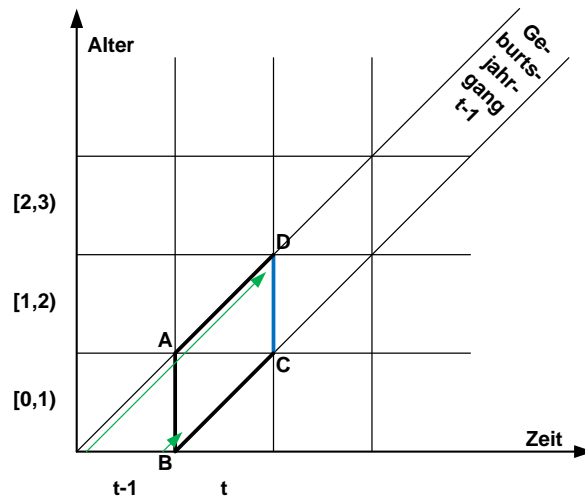


Abbildung B.2.: Bestimmung eines Alters für die Zu- und Fortzüge nach der amtlichen Statistik: Zu- und Fortzüge des Geburtsjahrgangs  $t - 1$  werden dem Altersintervall  $[1, 2)$  im Beobachtungsjahr  $t$  zugeordnet (blaue Linie); jedoch findet Migration tatsächlich im gesamten Parallelogramm  $ABCD$  im Beobachtungsjahr  $t$  statt, sodass Migranten des Geburtsjahrgangs  $t - 1$  zwischen einem Tag und fast zwei Jahren alt sein können (siehe grüne Linien).

werden.

Neben den Geburten und den Zu- und Fortzügen muss auch die Mortalität zur Berechnung der exakten Jahresdurchschnitts- und Jahresendbestände der zu prognostizierenden Bevölkerung betrachtet werden. Da die Geburten zur Jahresmitte und die Zu- und Fortzüge zum Jahresanfang und zum Jahresende stattfinden, wird jedes Beobachtungsjahr bzw. jedes Projektionsintervall in zwei Hälften geteilt. Jedes Halbjahr wird dabei durch die Jahresdurchschnittsbestände und die Jahresendbestände begrenzt. Um nun einen Bevölkerungsbestand zur Jahresmitte oder zum Jahresende zu berechnen, müssen der entsprechende Bevölkerungsbestand vor einem halben Jahr und die in diesem Halbjahr stattfindenden demographischen Ereignisse der Migration, Mortalität und Fertilität berücksichtigt werden. Folglich wird nicht nur die Migration, sondern auch die Mortalität aufgespalten, sodass für jedes Halbjahr bestimmte Überlebensverhältnisse im PPPM angenommen werden können. So werden für jedes ganze Altersintervall,  $[x, x + 1)$ , zwei Überlebenswahrscheinlichkeiten aus den anzunehmenden Werten für den Modellparameter *Überlebende im Alter  $x$*  (vgl. Abschnitt 10.1.1 auf Seite 187) gebildet. Die erste Überlebenswahrscheinlichkeit  $p_1$  betrifft

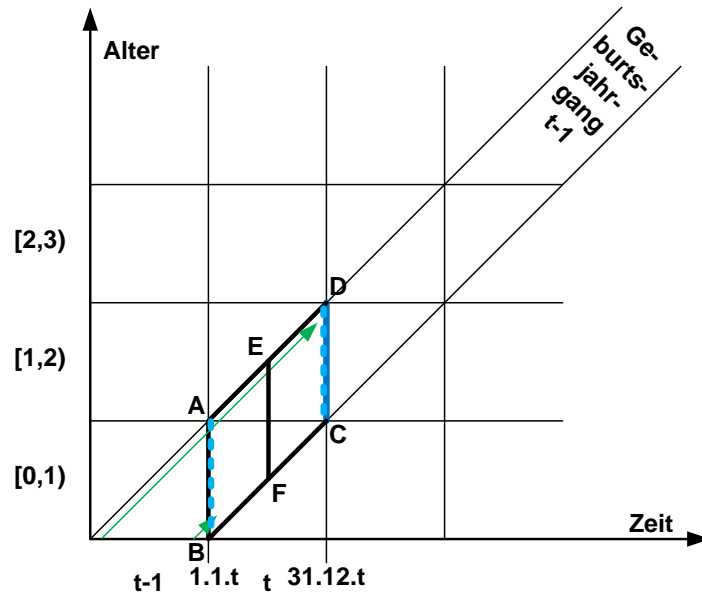


Abbildung B.3.: Festlegung der Zeitpunkte für die Migration im PPPM: Migration des Geburtsjahrgangs  $t - 1$  findet jeweils zur Hälfte am Beginn und am Ende (hellblau gestrichelte Linien) des Beobachtungsjahrs  $t$  statt; damit wird die Verteilung der Migranten im gesamten zu betrachtenden Altersintervall näherungsweise eingefangen.

das erste halbe Altersintervall  $[x, 0; x, 5)$  und die zweite Überlebenswahrscheinlichkeit  $p_2$  betrifft das zweite halbe Altersintervall  $[x, 5; x + 1)$ . Da die Jahresdurchschnittsbestände im Durchschnitt im Alter  $x, 0$  und die Jahresend- bzw. die Jahresanfangsbestände im Durchschnitt  $x, 5$  Jahre alt sind, wird die erste Überlebenswahrscheinlichkeit,  $p_1$ , verwendet, um von einem Jahresdurchschnittsbestand den nächstfolgenden Jahresend- bzw. Jahresanfangsbestand zu berechnen. Die zweite Überlebenswahrscheinlichkeit,  $p_2$ , wird demzufolge verwendet, um von einem Jahresend- oder Jahresanfangsbestand den nächstfolgenden Jahresdurchschnittsbestand zu berechnen (vgl. Abbildung B.4).

Allerdings ist eine einfache Aufspaltung der Mortalität aufgrund der Erfassung der Sterblichkeit über die amtliche Statistik nur mit einer Aufschiebung bzw. Versetzung um ein halbes Jahr möglich. Der Sterbetafelparameter *Überlebende im Alter  $x$*  wird von der amtlichen Statistik für ein Altersjahr  $[x, x + 1)$  in einem Kalenderjahr  $t$  aus den tatsächlich beobachteten Sterblichkeitsdaten einer Bevölkerung berechnet. Dabei erstrecken sich diese  $l_x$  vom 1.1. $t$  bis zum 31.12. $t$ . Im PPPM erstrecken sich diese  $l_x$  hingegen vom

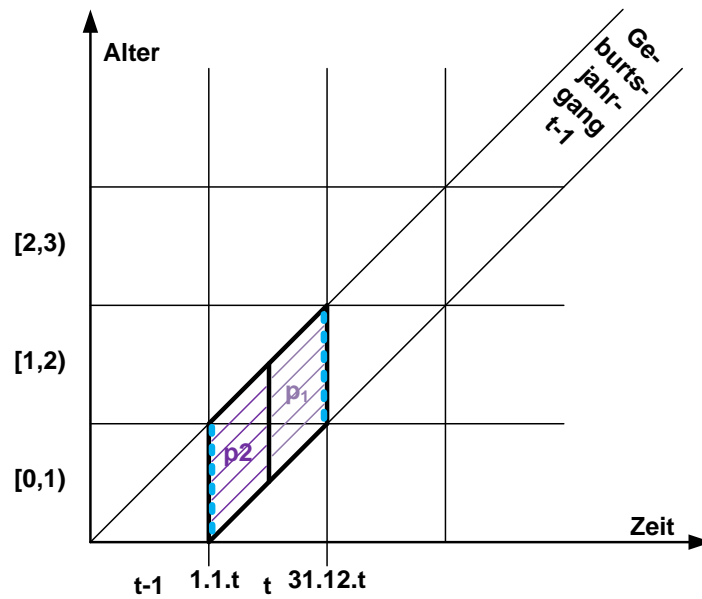


Abbildung B.4.: Festlegung der Zeitpunkte für die Migration und die Mortalität im PPPM am Beispiel des Geburtsjahrgangs  $t - 1$  im Beobachtungsjahr  $t$ : Demographische Ereignisse der Migration und Mortalität finden in Halbjahren statt; Migration findet jeweils zur Hälfte am Beginn und am Ende (hellblau gestrichelte Linien) des Beobachtungsjahrs  $t$  statt; Mortalität findet mit der ersten Überlebenswahrscheinlichkeit  $p_1$  (helllila) vom Jahresdurchschnittsbestand zum Jahresendbestand und mit der zweiten Überlebenswahrscheinlichkeit  $p_2$  (lila) vom Jahresanfangsbestand zum Jahresdurchschnittsbestand statt.

30.6. $t$  bis zum 30.6. $t + 1$  aufgrund der Annahme, dass Geburten immer nur zur Jahresmitte stattfinden. Folglich müssen die aus den Daten der amtlichen Statistik berechneten halbjährigen Überlebenswahrscheinlichkeiten,  $p_1$  und  $p_2$ , vom 1.1. $t$  bis zum 31.12. $t$  um ein halbes Jahr weiter aufgeschoben werden, sodass sie nun für das Jahr vom 30.6. $t$  bis 30.6. $t + 1$  gelten. Demzufolge ist  $p_{1,t}$  mit den Daten der amtlichen Statistik für das Jahr  $t$  berechnet worden und gilt im PPPM für die zweite Jahreshälfte des Jahres  $t$ . Äquivalent ist  $p_{2,t}$  ebenfalls mit den Daten der amtlichen Statistik für das Jahr  $t$  berechnet worden, gilt aber im PPPM für die erste Jahreshälfte des Jahres  $t + 1$  (vgl. Abbildungen B.5 und B.6).

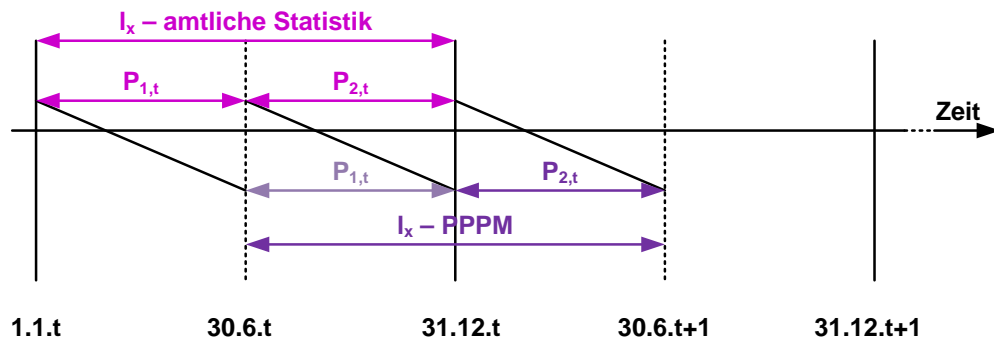


Abbildung B.5.: Verschiebung des Gültigkeitsbereichs der halbjährigen Überlebenswahrscheinlichkeiten  $p_{1,t}$  und  $p_{2,t}$  im PPPM aufgrund der Annahme, dass Geburten nur zur Jahresmitte stattfinden; diese Abbildung basiert (dankenswerterweise) nach vorheriger mündlicher Absprache auf einer Abbildung aus den Übungsmaterialien zur Vorlesung Migration II von Thomas Salzmann.

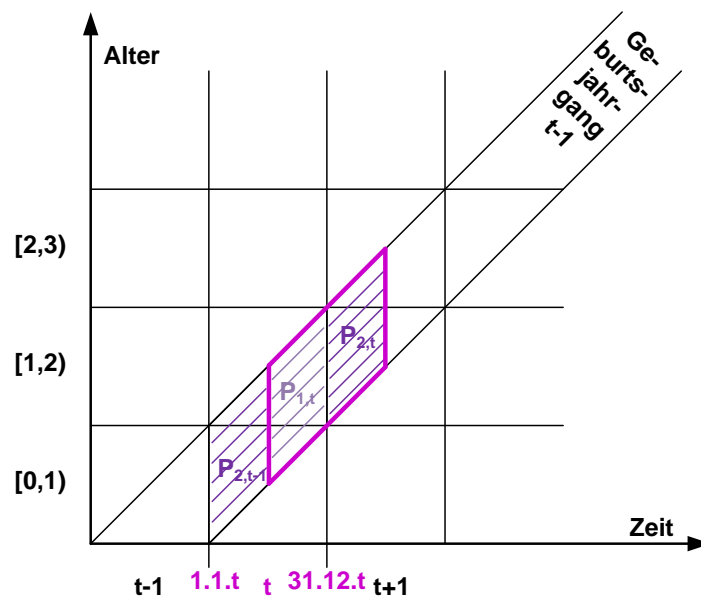


Abbildung B.6.: Tatsächlicher Gültigkeitsbereich der halbjährigen Überlebenswahrscheinlichkeiten  $p_{1,t}$  und  $p_{2,t}$  im PPPM:  $p_{1,t}$  gilt für die zweite Hälfte des Jahres  $t$  und  $p_{2,t}$  gilt für die erste Hälfte des Jahres  $t+1$ ; folglich ist der Gültigkeitsbereich der beiden Überlebenswahrscheinlichkeiten des Jahres  $t$  um ein halbes Jahr aufgeschoben (im Beerenton umrandetes Parallelogramm).

## C. *Repräsentative* Annahmenverläufe

An dieser Stelle wird die stochastische Variation eines repräsentativen Annahmenverlaufs illustriert.

In den Abschnitten 10.2.1 und 10.4.4 sind die repräsentativen Annahmenverläufe formal erläutert worden; im Kapitel 12 sind sie dann in den Varianten III und IV in der Beispielprojektion praktisch angewendet worden.

Die Abbildung C.1 zeigt nun eine stochastische Variation des blauen repräsentativen Annahmenverlaufs zur Sterblichkeit der einheimischen Männer aus der Beispielprojektion genauer; dabei ist zu erkennen, dass die repräsentative lineare Veränderung der Sterblichkeit über die Zeit in der stochastischen Variation durch zufällige Schwankungen aufgebrochen wird, sodass das Niveau wie auch die zeitlichen Abstände zwischen Niveauveränderungen zufällig variieren.

Je höher die Standardabweichung  $\sigma$  bestimmt wird, desto größer sind die potentiellen stochastischen Variationen der Annahmenwerte der jeweiligen Modellparameter.

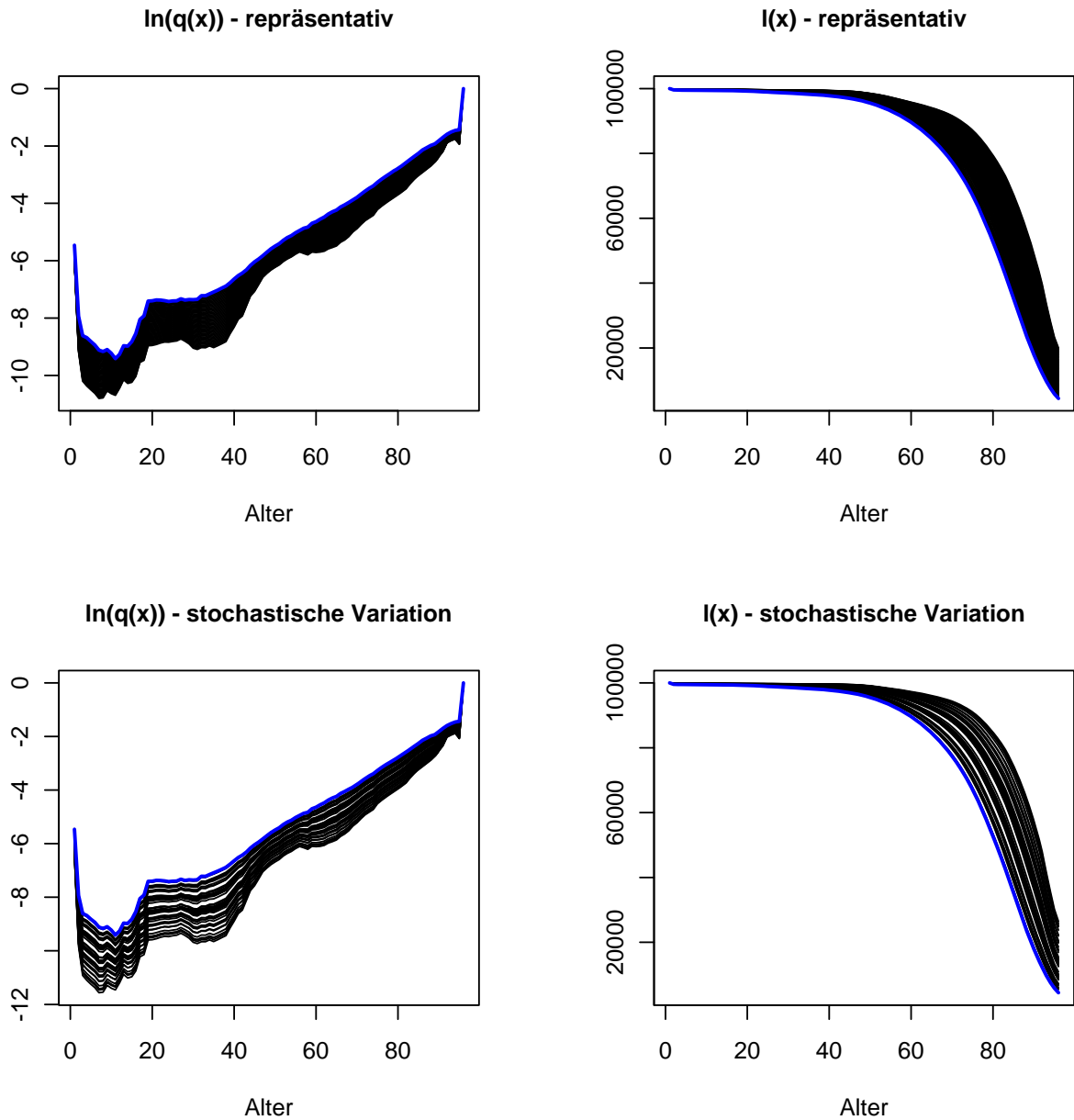


Abbildung C.1.: Exemplarische stochastische Variation des blauen repräsentativen Annahmenverlaufs zur Sterblichkeit der einheimischen Männer in den Varianten III und IV der Beispielprojektion aus dem Kapitel 12; die blaue Linie entspricht jeweils den Werten aus dem Jump-off Jahr.